



44 25 1.5 447 109 R40271 yf80f04.s1 Homo sapie 2.48e-06  
45 24 1.5 450 216 W13185 ma81d02.r1 Soares mou 4.42e-05

ALIGNMENTS

RESULT 1  
LOCUS T94579 428 bp mRNA EST 24-MAR-1995  
DEFINITION yef1905.s1 Homo sapiens cDNA clone 119384 3'.  
ACCESSION T94579  
NID 9728067  
KEYWORDS EST.  
SOURCE human clone-119384 library-Stratagene lung (#937210)  
vector-pBluescript SK- host-SOLR cells (kanamycin resistant)  
primer--21ml3 Rsttel-EcoRI Rsttel2-XhoI Normal lung tissue from a 72  
year old male. Cloned unidirectionally. Primer: Oligo dt. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:  
5'-GAATTCGGCAGAG-3'; 3' adaptor sequence:  
5'-CTCGAGTTTCTTTTCTTTT-3'.

ORGANISM  
Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 428)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,  
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 241  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
source  
1..428  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="119384"

BASE COUNT 112 a 103 c 124 g 76 t 13 others

ORIGIN  
Query Match 21.3%; Score 348; DB 151; Length 428;  
Best Local Similarity 94.4%; Pred. No. 0.00e+00;  
Matches 386; Conservative 0; Mismatches 18; Indels 5; Gaps 5;

Db 4 gaagaggacacaaacaaagatttttttgcagtgaaaggagcc-aggagaccga 62  
|||||  
Cp 1624 GAAGAGGGGACAAACAAAGATTATTTCAGAGTGAAGGGGCGCCAGAGAGCCCA 1565  
|||||  
Db 63 gaaaaaagaagaagagcttgagagagaagaagccttgataaagaagacacagaagaagc 122  
|||||  
Cp 1564 GAAAAAAGGAAGGCAAGGCTGAGAGCAGAAAGCTTGATTAAGAGACCAAGAAAGGC 1505  
|||||  
Db 123 ctgcaggaagcagatgtcgcccccagggaaaaacccaagagagcgtgtagagtatcc 182  
|||||  
Cp 1504 CTGCAGAGACCAAGATTGCGGCCGCCAGGAAAAACCAAGAGGAGAGCTGTAGAGTATCC 1445  
|||||  
Db 183 tggggccagcctcaaaagctggaacttgagagggccttan-gaactcaattccaggtcagca 241  
|||||  
Cp 1444 TGGGCCCACTCAAGCTGCGGACTGAGAGGCGCTTTRAGGACTCAATTCAGAGTGCAGCA 1385  
|||||  
Db 242 ttgcagagantgtctgtaaacacagcgccttctcgagcagctt-ctgnaaacacmgcccgcn 300  
|||||  
Cp 1384 TTTCGAGAGGTTGCTGTAACAACAGGCGCTGTGCGGAGCTTGTCTGTAACAAGCGGCCCGC 1325  
|||||  
Db 301 tgcacaactgttagaagctggaacctcccgagagnttgtagatagagccatcaagctt-cg 359

|||||  
Cp 1324 TGCAACAGCTGTGAAGAGCTGGAACCGTTCCCGAGAGATTGGGATAGAGCCCATCAGCTTCC 1265  
|||||  
Db 360 ctgcagaagcgtttctcttgcagggcttngggcatgcttcaggttcaga 408  
|||||  
Cp 1264 CTGCAGAGACGTGTCTGTCTCCAGGCGCTGGGCCATGCT-CAGGTTCAGA 1217  
|||||

RESULT 2  
LOCUS T94272 405 bp mRNA EST 24-MAR-1995  
DEFINITION yef1905.r1 Homo sapiens cDNA clone 119384 5'.  
ACCESSION T94272  
NID 9727760  
KEYWORDS EST.  
SOURCE human clone-119384 library-Stratagene lung (#937210)  
vector-pBluescript SK- host-SOLR cells (kanamycin resistant)  
primer-M13RPI Rsttel-EcoRI Rsttel2-XhoI Normal lung tissue from a 72  
year old male. Cloned unidirectionally. Primer: Oligo dt. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:  
5'-GAATTCGGCAGAG-3'; 3' adaptor sequence:  
5'-CTCGAGTTTCTTTTCTTTT-3'.

ORGANISM  
Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 405)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,  
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 214  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
source  
1..405  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="119384"

BASE COUNT 64 a 113 c 102 g 115 t 11 others

ORIGIN  
Query Match 15.0%; Score 245; DB 151; Length 405;  
Best Local Similarity 91.6%; Pred. No. 0.00e+00;  
Matches 371; Conservative 0; Mismatches 14; Indels 20; Gaps 20;

Db 1 tctgcaactgacatgagcccccagcccttgacaagaacagcttctgcagggcaaaactgat 60  
|||||  
Cp 1217 TGTGAACCTGACAGACAGGCCCGCCAGCCCTTGACAAAGACAGCTTCTGCAAGGCCAAAGCTAT 1276  
|||||  
Db 61 gggcctctatcccaatcctctcggaacggtccagcttctcaacagctgtgcagcgggcg-ctg 119  
|||||  
Cp 1277 GGGCTCTATCCCAATTCCTCGGGAAGGTCAGAGCTCTACAGCTGTGCAAGCGGGCGGCTG 1336  
|||||  
Db 120 ttccgcaaaagtgtgccgacag-ccctggtgttcagcaactcctgaaatgtgtcacctgg 178  
|||||  
Cp 1337 TTCCAGCAAAAGCTGCCACAGAGCGCTGTGTTCACCAACTCTGTCAAAATGCTGCACCTGG 1396  
|||||  
Db 179 gaattgagtcgttaagaagcccccagctccagcttgcagcttgagggccaggaattcacctc 238  
|||||  
Cp 1397 -AATTGAGTCGTTAAAGCCCTTCAGTCCAGCTTTGAGCTGGGCCCAAGG-AT-CAGTTC 1453  
|||||  
Db 239 ttacagccttgccttccctctgggggttttcccttggggggccgcgaatctggnttcttgcagc 298  
|||||  
Cp 1454 T-ACAGCGT-GCCT-CTGGG-TTTTCCCTGGGGG-CCGCAATCTGG-CTCCT-GCAGGC 1506  
|||||

Db	299	cttttttggtttttctctttancccaagggttttttgcctcaagcttgcctttt	358
Oy	1507	ctttct-cttctt-ctttta-tccaagctttct-gctc-cagccttgccttctttt	1561
Db	359	ttnttggttctcttggttgccnttcatltingcaantaat	403
Oy	1562	ttctt-gggttccttggtt-gcccttttcaactt-gcaaaataaatt	1603
RESULT	3	standard; DNA; STS: 241 BP.	
ID	HSg124		
AC	G29124		
NI	g1592433		
DT	04-OCT-1996 (Rel. 49, Created)		
DE	05-OCT-1996 (Rel. 49, Last updated, Version 2)		
DE	human STS SHGC-17364.		
KW	primer; sequence tagged site; STS sequence.		
OS	Homo sapiens (human)		
OC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
OC	Theria; Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.		
RN	[1]		
RP	1-241		
RA	Myers R.M.;		
RT	;		
RL	Unpublished.		
CC	Contact: Richard M. Myers Stanford Human Genome Center (SHGC)		
CC	Stanford University School of Medicine Department of Genetics,		
CC	M-344, Stanford, CA 94305, USA Tel: 4157259687 Fax: 4157259689		
CC	Email: myers@shgc.stanford.edu Primer A: AAAGCCTGGATAAAGAGAGACC		
CC	Primer B: GCTGCACCTGGATATGACG STS size: 150 PCR Profile: Initial		
CC	Incubation: 94 degrees C for 90 seconds Denaturation: 94 degrees C		
CC	for 15 seconds Annealing: 62 degrees C for 23 seconds		
CC	Polymerization: 72 degrees C for 30 seconds PCR Cycles: 30 Thermal		
CC	Cycler: Perkin Elmer 9600 Protocol: Template: 25 ng Primer: each 1		
CC	ul dnmps: each 200 uM Taq Polymerase: 0.05 units/ul Total Vol: 10		
CC	ul Buffer: MgCl2: 2.5 mM KCl: 50 mM Tris-HCl: 20 mM pH: 8.3		
CC	Prepared with primer pairs derived from T94579--Merck/Unilest.		
FH	Key	Location/Qualifiers	
FT	source	1..241	
FT		/organism="Homo sapiens"	
FT		/map="1"	
FT	STS	91..240	
FT	primer_bind	91..112	
FT	primer_bind	complement(221..240)	
FT	Sequence 241 BP, 78 A; 53 C; 77 G; 30 T; 3 other:		
SO	Query Match	13.3%; Score 218; DB 250; Length 241;	
	Best Local Similarity 97.1%; Pred. No. 0.00e+00;		
	Matches 233; Conservative 0; Mismatches 5; Indels 2; Gaps 2		
Db	4	gaagangggcaacaacaaagaattattttgcaagtgaaggcgagcc-aggagacc	62
Cp	1624	GAAGAGGGGCGCAACCAACCAAGATTATTATTTCAGATGAAAGGGCGAGCCCGAGAGACCA	1565
Db	63	gaaaaaagaaggaagcaagcctgaagcgagaagcctgtgtaagaagaagacacaaagaagc	122
Cp	1564	GAAAAAGGAAGGCGAAGGCTGAGAGCAAAAGGCTGTGATTAAGAGAGCCACAAGAAAGGC	1505
Db	123	ctgcagagcgcaagatttggtggcccccaagggaacccacgaagagcgagctgttagatgacc	182
Cp	1504	CTGCAGAGCGCAAGATTGGGGCCCCCAGGAAACCCAGAGAGCGAGCGTGTGAGTGTCC	1445
Db	183	tgggcccagcctcaaaagcttggaacttggtggggccttan-gactcaattcgaagtgcagca	241
Cp	1444	TGGGCCCAAGCCTCAAAAGCTGGAGACTGGAGGGGGCTTTAAGCAGCACTCAATTCCAGGTGACAGCA	1385
RESULT	4	G29124	241 bp
LOCUS	DEFINITION	human STS SHGC-17364.	DNA
ACCESSION	G29124		STS
			04-OCT-1996

**NID** 1592433  
**KEYWORDS** STS sequence; primer; sequence tagged site.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 241)  
**AUTHORS** Myers, R.M.  
**JOURNAL** Unpublished (1996)  
**COMMENT**  
 Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 415/7259687  
 Fax: 415/7259689  
 Email: myers@shgc.stanford.edu  
 Primer A: AAGCCTGGATTAAGGAAGAC  
 Primer B: GCTGCACCTGGATTGATC  
 STS size: 150  
 PCR Profile:  
 Initial incubation: 94 degrees C for 90 seconds  
 Denaturation: 94 degrees C for 15 seconds  
 Annealing: 62 degrees C for 23 seconds  
 Polymerization: 72 degrees C for 30 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600  
 Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Tag Polymerase: 0.05 units/ul  
 Total Vol: 10 ul  
 Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 20 mM  
 pH: 8.3  
 Prepared with primer pairs derived from T94579-Merck/UniEST.  
 Location/Qualifiers  
 1..241  
 /organism="Homo sapiens"  
 /map="1"  
 STS  
 primer\_bind 91..240  
 91..112  
 primer\_bind complement(221..240)  
 BASE COUNT 78 a 53 c 77 g 30 t 3 others  
 ORIGIN  
 Query Match 13.3%; Score 218; DB 202; Length 241;  
 Best Local Similarity 97.1%; Fred. NO. 0.00e+00;  
 Matches 233; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
 Db 4 gaagngggcacaacccaagattatttttcgaagtgaaggagcagcc-aggagaccga 62  
 Cp 1624 GAAGGGGGGCACAAACCAATTATTATTTCAGTAGAAAGGCGACGCCAGGAGACCCA 1565  
 Db 63 gaaaaaagaaggaagcaaggtctagagcagaagaagcctgtgtaaggaagacacacagaagaagc 122  
 Cp 1564 GAAAAAAGAGAGGCGACAGGCTGAGAGCACAAGAGCCTGTGATTAAAGAGACACACAGAAAGCC 1505  
 Db 123 ctgcagagcagcagattgtcgcccccagaggaaaaccacagagagcagcgtgtagagtgatcc 182  
 Cp 1504 CTGCGAGGACCAAGATTGGCGGCCCCAGGAGAAACCCAGAGGAGCGCTGTAGAGTGATCC 1445  
 Db 183 tggggccagcctcaaaagtgtgactgtggggggcctttaa-gactaatctccagggcgagca 241  
 Cp 1444 TGGGCCCGCTCAAAAGCTGGAGACTGGAGGGGGCTTTAGCGGACTCAATTCCAGGTGCAGCA 1385

[illegible][illegible]



QY 199 GAGCACCAGCTGAGTGGATGACAGACTCTCTACAGAGATTCAATGCGCTGAAGAAGAT 258  
 Db 266 gaatcccaa 274  
 QY 259 GAATCCCAA 267

RESULT 7  
 LOCUS N75653 541 bp mRNA EST 02-APR-1996  
 DEFINITION yv29c01.r1 Homo sapiens cDNA clone 244128 5' similar to gp:M80927  
 CARRIER GLYCOPROTEIN-39 PRECURSOR (HUMAN).  
 ACCESSION N75653  
 NID 91238231  
 KEYWORDS EST.  
 SOURCE human clone=244128 primer=reverse ET library=Soares fetal liver spleen INFs vector=PT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) RstIel-Pac I RstIe2-Eco RI liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - 01590(dT) primer [5].  
 AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'1, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens  
 Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrabrata; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 541)  
 Hiller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
 The Washu-Merk EST Project  
 Unpublished (1995)

TITLE The Washu-Merk EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Wilson RK  
 Washu-Merk EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 441  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

NCBI g1: 1238231  
 FEATURES  
 source location/Qualifiers  
 1..541  
 /organism="Homo sapiens"  
 /clone="244128"  
 /note="human"  
 1..>541  
 BASE COUNT 149 a 134 c 117 g 136 t 5 others  
 ORIGIN

Query Match 4.88; Score 78; DB 207; Length 541;  
 Best Local Similarity 66.7%; Pred. No. 8.69e-99;  
 Matches 166; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

Db 27 ctggagaggtgtgtgtctgtctgtctccagggagagatgcctcaaaactgtttg 86  
 QY 19 ctggcaggggtttatgctctctgatgatccatggggcttgcgcaaaactggctg 78  
 Db 87 ctacttaccactgtgtccagcagccgagaaacgaagaaa-ctcaacccctgagaatat 145  
 QY 79 ctacttcacacactggcggccagatgacagacaggggagggcctgccttccgtcccaagact 138

Db 146 tgacccttcctatgtctcatcctatcattcattcgcgcacatcgaaacaaaggt 205  
 QY 139 GGACCCAGCCTTTTGACACCACTCATCTACGCTTCGTGGCATACCAACACCAAGCT 198  
 Db 206 tatcatcaagaagaaagtgatgtgtctctacagacatcaacagtctcaaaccaa 265  
 QY 199 GAGCACCAGTGAAGTGAATGACAGACTCTCTACAGAGATTCAATGCGCTGAAGAAGAT 258  
 Db 266 gaatcccaa 274  
 QY 259 GAATCCCAA 267

RESULT 8  
 LOCUS T65854 533 bp mRNA EST 20-FEB-1995  
 DEFINITION yv24b03.s1 Homo sapiens cDNA clone 81581 3' similar to SP:S27879  
 S27879 SECRETORY PROTEIN YN-1 PRECURSOR -.  
 ACCESSION T65854  
 NID 9674899  
 KEYWORDS EST.  
 SOURCE human clone=81581 library=Stratagene lung (#937210)  
 vector=PB1uescript SK- host=SOLR cells (kanamycin resistant)  
 primer=-21m13 RstIel-EcoRI RstIe2-XhoI Normal lung tissue from a 72 year old male. Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGGCAGCAG-3'; 3' adaptor sequence: 5'-CTCGAGTCTTTTCTTTTCTTTT-3'.

ORGANISM Homo sapiens  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 533)  
 Hiller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
 Washu-Merk EST Project  
 Unpublished (1995)

TITLE Washu-Merk EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Wilson RK  
 Washu-Merk EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 307  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES  
 source location/Qualifiers  
 1..533  
 /organism="Homo sapiens"  
 /clone="81581"  
 BASE COUNT 140 a 123 c 136 g 128 t 6 others  
 ORIGIN

Query Match 4.5%; Score 73; DB 144; Length 533;  
 Best Local Similarity 70.1%; Pred. No. 5.24e-89;  
 Matches 131; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

Db 250 tggacaagaggtgaagagccctgtgtgcagagatttccagtgtaagtcacatgttaata 309  
 Cp 1137 TGGATAGGGGGGTATGGGCTGTGTGCGAGAAAGCCGGCAAGTCAATGATCACT 1078  
 Db 310 gaccagatcatgctctccccaaggttaattcttgaagaactgaacactgtgtctcata 369  
 Cp 1077 GCCAACAATGCCCCGCCCAATCCCTTCTGTTAGATAGCTGACCTTGGTTGAAG 1018  
 Db 370 ctcttcacatcatatagaccacacactgtgttcccttgactnctgaagggagcctctg 429  
 Cp 1017 CTCTCACAATCANCAAGCCACCACT-GGTTGCTCCGGAAGATGTAGGGACCTTCTG 959  
 Db 430 atctctgg 436

Cp	958	ATCCGTGG	952	!
RESULT	9			
LOCUS	G24226	533 bp	DNA	STS
DEFINITION	human STS WI-14165.			31-MAY-1996
ACCESSION	G24226			
NID	9134452			
KEYWORDS	STS sequence; primer; sequence tagged site.			
SOURCE	human STS derived from sequences in dbEST and the Unigene collection.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Carnivora; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 533)			
TITLE	Hudson, T.			
JOURNAL	Whitehead Institute/MIT Center for Genome Research; Physically mapped STS			
COMMENT	Unpublished (1995)			
	Contact: Thomas Hudson			
	Whitehead Institute/MIT Center for Genome Research			
	Whitehead Institute for Biomedical Research			
	9 Cambridge Center, Cambridge MA 02142 USA			
	Tel: 617 252 1900			
	Fax: 617 252 1902			
	Email: thudson@genome.wi.mit.edu			
	Primer A: AGGGCAACAATCTCAACAGCT			
	Primer B: TCTCCGAGATTCTCATGTGG			
	STS size: 132			
	PCR Profile:			
	Presoak:			
	Denaturation:			
	Annealing: 56 degrees C			
	Polymerization:			
	PCR Cycles: 35			
	Thermal Cycler:			
	Protocol:			
	Template: 10 ng			
	Primer: each 5 pM			
	dNTPs: each 4 mM			
	Tag Polymerase: 0.025 units/ul			
	Total Vol: 20 ul			
	Buffer:			
	MgCl2: 1.5 mM			
	KCl: 50 mM			
	Tris-HCl: 10 mM			
	pH: 9.3			
FEATURES	Derived from dbEST (genbank accession T65854).			
source	location/Qualifiers			
1..533				
/organism="Homo sapiens"				
/map="461.5 CR from top of Chr1 linkage group"				
STS	34..165			
primer_bind	34..53			
primer_bind	complement(145..165)			
BASE COUNT	140 a 123 c 136 g 128 t 6 others			
ORIGIN				
Query Match	4.5%; Score 73; DB 200; Length 533;			
Best Local Similarity	70.1%; Pred. No. 5.24e-89;			
Matches	131; Conservative 0; Mismatches 55; Indels 1; Gaps 1;			
Db	250	tgacaagaaggttaagagcctctgttgcagagattncacagtgaagtcacatcatgcaata	309	
Cp	1137	TGATGTAAGGGGTATTCGCCCTCGTGTGCAAGAGAACCGCGCAAGTCATCTAATGCAGT	1078	
Db	310	gaccagatcatgtgctctctccacagtttaaatctttaagaactgaacctgtgtctcaca	369	

Cp	1077	GCACAGACATGGCCCGCCGACATCCCTCTGCTTCAGATACCTGACTGCTTGGTTTGAAG	1018
Db	370	ctcttcacatcatatagccaccacactggttcccttgactcncgtagggagactgctg	429
Cp	1017	CTCTCCACATCATCAAGCCACCCACT-GGTTGTCGCCGAAGATGTAGGCGACCTCTG	959
Db	430	atctctg 436	
Cp	958	ATCTCG 952	
RESULT	10		
LOCUS	NA0107	311 bp	EST
DEFINITION	257209.r1 Homo sapiens cDNA clone 257753.5, similar to gb:M80927		
ACCESSION	NA0107		
NID	91163652		
KEYWORDS	EST.		
SOURCE	human clone=257753 primer=77 library=Soares placenta 8to9weeks 2bBH8to9w vector=PT73 (pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI two placentae: one from 8 weeks and another from 9 weeks post conception. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATGTGAAGTGGGCGCGCGGATTTTATTTT-3'] double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified PT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaudo.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostoma; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Channata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 311)		
REFERENCE	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Hollman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.		
AUTHORS	The WashU-Merck EST Project		
TITLE	Unpublished (1995)		
JOURNAL			
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 268 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL IMAGE Consortium (info@image.llnl.gov) for further information.		
FEATURES			
Source	1..311 /organism="Homo sapiens" /clone="257753"		
BASE COUNT	93 a 81 c 68 g 68 t 1 others		
ORIGIN			
Query Match	4.3%; Score 71; DB 79; Length 311; Best Local Similarity 65.8%; Pred. No. 3,96e-85;		
Matches	156; Conservative 0; Mismatches 80; Indels 1; Gaps 1;		
Db	69 ctggagcagtgtagtggcttgctgctcttcacaggagagatgacctacaacgtgttg	128	
Qy	19 CTGGCAGGTTTCATGTCTCTCTGTGATGATCCCATGGGCTGTGCGCAAACTGGTCTG	78	
Db	129 ctactttacaacatggtgccagagccggagaaaccagsgaaaa-ttcaacctcgagaatat	187	
Qy	79 CTACTTACCAACTGGGCGCCAGTACAGACGGGGGAGGCTCGCTCTCTCCCAAGACTT	138	








RESULT	13	R55530	396 bp	mRNA	EST	22-MAY-1995
LOCUS		yj9gc12.r1 Homo sapiens cDNA clone 154966 5'			similar to gb:M80927	
DEFINITION		CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.				
ACCESSION		R55530				
NID		9824825				
KEYWORDS		EST.				
SOURCE		human clone=154966 library=Soares breast 2NBHst vector-pt7n3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13Rp1 Rsite1-Not I Rsite2-Eco RI Adult female. 1st strand cDNA was primed with a Not I - 01190(dT) primer [5' TGTTACCAGACTGATGGAGGAGCGCCGCCCTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Ornithota; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 396) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kueba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,B., Rohlfing,T., Soares,M., Tan,F., Trivaskis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.				
AUTHORS		The Mashu-Merck EST Project				
TITLE		Unpublished (1995)				
JOURNAL						
COMMENT		Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 284 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information.				
FEATURES		Location/Qualifiers				
Source		1..396				
		/organism="Homo sapiens"				
		/clone="154966"				
BASE COUNT		88 a 106 c 109 g 87 t 6 others				
ORIGIN						
Query Match		4.0%; Score 65; DB 134; Length 396;				
Best local Similarity		68.3%; Pwd. No. 1,28e-73;				
Matches		114; Conservative 0; Mismatches 53; Indels 0; Gaps 0;				
Db		125 cagcgttgatgtctcgtgagtgcgtcccaagtgtgtctgcacatacaaacgtgcttact 184				
Oy		24 CAGGTTCATCATGCTCCTCGATGATGCATCCATGGGCTCTCTGCATAAATCGTCTGCTACT 83				
Db		185 acacacagctggtcccaagtaaccgganngacatgtagtgagtcgttcccagatgcccttgacc 244				
Oy		84 TCACCAACTGGCGCCAGTCACAGACAGGGGAGAGGCTTCCTCGCCCAAGACTTGAGACC 143				
Db		245 gnnctctctgtaccacatcatcatcaagctttgccaatataagaac 291				
Oy		144 CCAGCCTTTGGCACCCACCTCATCTAGCCTTGCTGGCGCATGACCAAC 190				
LOCUS	14	M10705	306 bp	mRNA	EST	26-APR-1996
DEFINITION		ma47g07.r1 Soares mouse p3JMR19.5 Mus musculus cDNA 5'			similar to	
ACCESSION		gb:M80927			CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.	
		M10705				

NID	91285179
KEYWORDS	EST.
SOURCE	Mus mouse.
ORGANISM	Mus musculus Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrate; Eutheria; Rodentia; Sciurognathi; Mymorpha; Muridae;
REFERENCE	Murine; Mus. 1 (bases 1 to 306)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The Mashu-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Seq primer: mob.REGA+ET High quality sequence stop: 172.
FEATURES	NCBI g1: 1285179
source	Location/Qualifiers 1..306 /organism="Mus musculus" /note="Vector: pRTD3 (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTAGTCGAATGGCGGCCGCATTGTCTTTTTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTD3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. library constructed by Bento Soares and M.Fatima Bonaldo." /clone_lib="Soares mouse pJNMF19.5" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)" <1...>306
BASE COUNT	78 a 76 c 85 g 67 t
ORIGIN	mRNA
Query Match	3.7%; Score 60; DB 215; Length 306;
Best Local Similarity	72.1%; Pred. No. 3.39e-64;
Matches	98; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
Db	78 agaagcttcocctcgtcctaacaaggcaaccacggtgggtgatgaagacaagaagtgc 137
Oy	960 AGAAGGTGCGCTACACTCTTCGCGGACAACAAGTGGTGGCTTTATGATGTGGAGAGCT 1019
Db	138 tcaaaacaagaatttggttccctgaagaggaagaagaactgcgaagagcatcggtgtggcac 197
Oy	1020 TCAAAACCAAGTCAAGTCATCTCAGAAGCAAGAGGAGCTGGCGGGCCCATGCTGTGGCAC 1079
Db	198 tggatttgatgatctt 213
Oy	1080 TGGACTTAGATGACTT 1095
RESULT	15
LOCUS	W10705 306 bp mRNA EST 05-SEP-1996
DEFINITION	ma47q07.r1 Soares mouse pJNMF19.5 Mus musculus cDNA clone 313884 5'
ACCESSION	W10705 similar to gp:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.
NID	91285179
KEYWORDS	EST.









```

RESULT 2
LOCUS HUMHA3G 1741 bp mRNA PRI 24-JAN-1994
DEFINITION Human glycoprotein mRNA, complete cds.
ACCESSION M80927
NID 9348911
KEYWORDS glycoprotein.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1741)
Hakala,B.E., White,C., and Recklies,A.D.
Human cartilage gp-39, a major secretory product of articular
chondrocytes and synovial cells, is a mammalian member of a
chitinase protein family
JOURNAL J. Biol. Chem. 268 (34), 25803-25810 (1993)
MEDLINE 94064658
FEATURES
source 1..1741
/organism="Homo sapiens"
/cell_type="chondrocyte"
/tissue_type="cartilage"
1..71
5'UTR 72..134
sig_peptide 72..134
/codon_start=1
72..1223
CDS /note="articular 39kDa glycoprotein"
/codon_start=1
/evidence="experimental"
/produced="glycoprotein"
/db_xref="pid:9348912"
/translation="MGVRSQSGFVYLVLQCSAYKLVYYSWQYRBDGSCFPD
ALDRLCRTHIYSEFANISNDHIDTENDVLTIGMLTNKRNPMKTLISYGVNNG
SQRPSKIASNTQSRFTIKSVPELRTGFGDLAWLPGRNDQHPFTLIKEMKAE
FIRKQAPGRKQDLISALASAGVITDSSYDIASKLSQHLDFISIMYDHWGMRGTGH
HSPLEFGQEDASPRFSNTDYGVMRLGAPASLVAGIPFGSRFTLASSEGVGA
PISGPIRGFTKREAGTLAYEICDFELGATVHRITLGOVPATRGNGMWGDDQESV
KSNQYIKDRQLAGMVMALDLDPGSSFCQDLRFPLTKAKDLAAT"
mat_peptide 135..1220
/codon_start=1
/produced="glycoprotein"
3'UTR 1224..1741
polyA_signal 1715..1720
polyA_site 1741
BASE COUNT 416 a 510 c 447 g 368 t
ORIGIN
Query Match 11.2%; Score 184; DB 76; Length 1741;
Best Local Similarity 63.8%; Pred. No. 3.57e-154;
Matches 701; Conservative 0; Mismatches 373; Indels 24; Gaps 10;

```

```

QY 324 ATATGTAGCCACGGCCAAACCGTACAGCTTTGTCAACTCGGGCATCAGATTTCGC 383
DB 454 gcaccatggtcttcatggtctgacattgctgtcttaacctgga---c---ggag-- 505
QY 384 GCMAATACAGCTTTACGGCCTTGACCTTGAGTGGAGTACCCAGGAGGCCAGGCC 443
DB 506 -----agacaaagacatttacacctatcaaggaatgaaggcgaattataa 558
QY 444 CTGCGGTAGACACAGACCCCTTCACACCTTGTCACAGACTTTGGCAATGCTTCCACC 503
DB 559 aggaagccag--c--caggaaaaagacgctcctctcagcgcgaacactgtctcgggga 615
QY 504 AGGAAGCCAGACCTCAGGAGGAAGAACGCTTCTCTGTAGTACACGGTTCAGCGTGGCC 563
DB 616 aggtacacattgacagcagctatgacattgccaatatccacaacctgtattcaata 675
QY 564 AGACCTATGTGATCTGTGATCGAGGTGACAAATCCCGAAGACCTGGATTTTGTCA 623
DB 676 gcatcatgacctagatttcatgttgagcctgctggtgggacacagggccatcacatcccc 735
QY 624 ACCTTATGCGCTTACGACTTCCATGGCTCTTGGAGAGAGTTCAGGAGACTTAACAGCCCC 683
DB 736 tgtccgaagtcaggagagatgcaagtccttgacagattccagacacatgactatgtgtg 795
QY 684 TCTACAGAGGCGAAGAAGAGAGTGTGACAGCAGCCAGCCTCAACGTGATGCTGCTGTC 743
DB 796 gttacatgttagtggtgggtcctctgcagtaagctgtgtatgttgatccccacctcg 855
QY 744 AACAGTGGCTCGAAGAGGGGACCCTGCGACGACAGCTGATCCTTGATGCTTACTACG 803
DB 856 ggaagagctcctctctgtctctc-tgagactg--ggtgtgaagcccaactcagagc 912
QY 804 GACGCTCTTCACTGCTGCTCTCTCATCAGACACAGAGTGGGGCCCGCCAGCACAGGGT 863
DB 913 cgggaattccagccggttccacaagagagcagggacccttccactactatgagatctgt 972
QY 864 CTGGCACTCCAGGCCCTTCCACCAAGAGAGAGGATGCTGGCTCACTTGAAGTCTGT 923
DB 973 acttcctccgcgagacacagatccatagaacctcggccaagagtcctcctatgcacaa 1032
QY 924 CTTGGAAG-G-GG-GCCACCAACAGAGATCCAGAGTGAAGAGTGCCTCACTTTC 980
DB 1033 agggcaaccagtggttagatcagcagcaggaagcgtcaaaagcgaagtgcagtaac 1092
QY 981 GGGACAAACGAGTGGGTGGCTTGTGATGTGAGAGACTTCAAAACCAAGCAGACTATC 1040
DB 1093 tgaagatagcagcttgagcagcgccatgtagttagggccctgagcctgagatccag 1152
QY 1041 TGAAGCAGAGGAGACTGGGCGGGCCATGCTGTGGGCACTGACATTAGATGACTTGGCCG 1100
DB 1153 gctcctctgcggccag 1170
QY 1101 GCTTCTCTGCAACACAGG 1118

```

```

RESULT 3
LOCUS HSU49835 1418 bp mRNA PRI 25-JUL-1996
DEFINITION Human YKL-39 precursor mRNA, complete cds.
ACCESSION U49835
NID 91457940
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1418)
Hu,B., Trinh,K., Figueira,W.F. and Price,P.A.
Isolation and sequence of a novel human chondrocyte protein related
to mammalian members of the chitinase protein family
JOURNAL J. Biol. Chem. 271 (32), 19415-19420 (1996)
MEDLINE 96325055
REFERENCE 2 (bases 1 to 1418)
AUTHORS Price,P.A.

```

TITLE Direct Submission  
JOURNAL Submitted (23-FEB-1996) Paul A. Price, Biology, 0322, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0322, USA

FEATURES Location/Qualifiers

source

1..1418  
/organism="Homo sapiens"  
/tissue\_type="articular cartilage"

CDS

36..1193  
/note="Similar to human glycoprotein encoded by Genbank Accession Number M80927, human chondroitinase precursor encoded by Genbank Accession Number U29615, human ovoductal glycoprotein encoded by Genbank Accession Number U09550, and to mouse secretory protein YM-1 precursor encoded by Genbank Accession Number S27879"

/codon\_start=1  
/product="YKL-39 precursor"  
/db\_xref="PID:91457941"  
/translation="MDQSLMAGVYVLLLLQGSAYKLVCFYTNMSQDRQPEKTPPE  
NIDPFLCSHLIYSPASILENNKVIKDKSEYMLYPTINSLTKNPKLILSIGYLFG  
SKGHPWVDSSTSLRLEFINSLIEFLRNHNDGLDVSIVYEDQKENTHEVLIHLAIA  
FQDEFTSTKRELITAGVAGROMDINSYOVERLAIDLDFINLSDFHGSMEKPII  
TGHNSPLSKMORBPSSSYNVEAVYWHKMPSSKRYMGPIYTHSPTLSAETT  
VGAPASCPGAPGPTTESGFLATIEICQFLKGAITLDDQVYAVKGMQWGYDDV  
KSMETKYQFLKNNLNGAMIMSIDMDFTGKSCNQGYPPIVOAVKRLSLGL"

sig\_peptide

36..98  
/product="YKL-39"

mat\_peptide

99..1190  
/product="YKL-39"

polyA\_site

1418  
/note="18 A nucleotides"

BASE COUNT

370 a 357 c 337 g 354 t

ORIGIN

9.7%: Score 159; DB 74; Length 1418;  
Best Local Similarity 63.0%; Pred. No. 5,026-118;  
Matches 711; Conservative 0; Mismatches 350; Indels 27; Gaps 10;

Query Match

53 ctggagcagtgtagtgcttgcttgcttcctcgaaggagagatcgctcctaactggttg 112

19 CTGGGAGAGTTTCAMGGTCTGCTGATGATCCCATGGGCTCTGTCGAAAACGTGTG 78

113 ctacttaccactggtcccgagaccgaggaaccaggaataatccaccctgagaatat 172

79 CTACTTCACCAACTGGGCGCCAGTACAGAGGAGGAGCTGCTCTGCCCAACGAGACTT 138

173 tgaccocctctatgctctctcctcctcctcctcctcctcctcctcctcctcctcct 232

139 GGACCCGAGCCTTTCACCCACCTCCTATCTACGCTTCCTGGCATGACCAACCACT 198

233 tatcatcaagacaagaagtgatgatctctaccagaccatcaacagcttccaaccaa 292

199 GAGCCACCACTGATGATGAGAGACACCTCTACCAAGAGTTCAATGGCTCGAAGAAAT 258

293 gaatcccaactgaataatctctcttgatcattgagaggtgactcgttggttccaaagg 352

259 GAATCCCAAGCTGAAGACCTCTGTAGCATGAGAGCTGGAATTTGGCAGCTCAAGATT 318

353 caaccctatggtgattcttctaccacgcttggaattcatttaactccaataatcctgt 412

319 CACAGATATGTGATGACCAAGGCAACACCGGACGCTTTGTGCACTGGCCATAGAGTT 378

413 tctgaggaacataacttgatgagctgagtgatgagtgatgagtgatgagtgatgagtg 469

439 GAGCCCTCCGTAGACAGAGAGCGCTTCACCAACCTGAGACGACTTGGCCAAATGCTT 498

470 -agaa-----acactcat--ttcacttgctgattcattagattagcagaaccc 517

439 GAGCCCTCCGTAGACAGAGAGCGCTTCACCAACCTGAGACGACTTGGCCAAATGCTT 498

518 tcagaagagacttcacaaataccacaaaggaaggcttctctgagctggggcgatctcgc 577

499 CCAGAGGAGAACCCAGAGCTCAGGAGGAGAACGCGCTTCTTGAAGTGAAGCGGCTTCAC 558

Db 578 agggagcacaatgattgataacagctatacaagtgaagaactcgcaaaagatcgtgatt 637

QY 559 TGGGACACACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 618

Db 638 catcaactctctgctcttgatcctcattcattcattcattcattcattcattcattcatt 697

QY 619 TGTCAACCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 672

Db 698 caaacgacctctgagcaaggagtgagcaaggagtgagcaaggagtgagcaaggagtgag 757

QY 673 TACAGCCCTCTTACAG 732

Db 758 atatgctgctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 817

QY 733 TGTGCTGCTGCAAG 792

Db 818 cccacataatgagcattcctcattcattcattcattcattcattcattcattcattcatt 874

QY 793 GCTTACCTACGAG 852

Db 875 tgcctcggccctgagcctgagcctgagcctgagcctgagcctgagcctgagcctgagc 934

QY 853 AGCCACAGGCTCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 912

Db 935 tgaatctgagcagtgagcagtgagcagtgagcagtgagcagtgagcagtgagcagtgag 994

QY 913 TGAAGTCTGC---TCTGAG 969

Db 995 ctacgagcagtgagcagtgagcagtgagcagtgagcagtgagcagtgagcagtgagcag 1054

QY 970 CTACATCTTCGAG 1029

Db 1055 gttcagcttcttaagaatttaaacctgagagagagagagagagagagagagagagag 1114

QY 1030 GGTACGATATGTAAG 1089

Db 1115 tgacttactgagcagtgagcagtgagcagtgagcagtgagcagtgagcagtgagcag 1162

QY 1090 TGACTTTGCGCGCTTCTGCTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137

RESULT 4

LOCUS HSU58514 1434 bp mRNA PRI 24-JUL-1996

DEFINITION Human chitinase precursor (HUMCHIT) mRNA, exon 1a form, complete

cds.

ACCESSION U58514

KEYWORDS 91439565

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1434)

Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1434)

Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

REFERENCE 2 (bases 1 to 1434)

The Washu-Merck EST Project

REFERENCE 3 (bases 1 to 1434)

Grossman, A., Matsuyama, T., Baker, E., Waterhouse, P., Sutherland, G. R., and Mak, T. W.

REFERENCE 4 (bases 1 to 1434)

Cloning of a novel lymphoid restricted human chitinase and localization to Ipl3.3

REFERENCE 5 (bases 1 to 1434)

Grossman, A., Matsuyama, T., Baker, E., Waterhouse, P., Sutherland, G. R., and Mak, T. W.

REFERENCE 6 (bases 1 to 1434)

Direct Submission

JOURNAL Submitted (17-MAY-1996) Medical Biophysics, Ontario Cancer

Institute, Princess Margaret Hospital, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

FEATURES  
SOURCE  
Location/Qualifiers  
1..1434  
/organism="Homo sapiens"  
/map="1p13.3"  
/chromosome="1"  
/tissue\_type="Soares placenta"  
/dev\_stage="8 to 9 weeks"  
/clone="EST 257753"  
1..36  
/gene="HUMTCHIT"  
1..76  
/gene="HUMTCHIT"  
/note="exon 1a; first alternatively splice exon 1; contains a kozak consensus initiator methionine"  
5'UTR  
1..114  
/number=1  
37..1209  
/gene="HUMTCHIT"  
37..1209  
/gene="HUMTCHIT"  
/note="first alternatively spliced form, see Genbank Accession Number U58515 for the second alternatively spliced form"  
/product="chitinase precursor"  
/db\_xref="pid:g1439566"  
/translation="MGATMDQKSLMAGVYVLLLOGGSNRYCYFTNMSQDREP  
KTPENIDPLCSLITFSASINNKVILDKSEVMIDYOTINSLKTKNPKLILLISG  
GYLFSGKGFPMVDSSISRLFINSLILFRNHPGLDYSWIPPOKNEHTVYI  
ELAEAFQKDEPTKSTKRLTLTAGVSAGROMIDNSYVEKLANKDLDFINLSPFHS  
EKPLITGNHPLSLKGMODRGPSSYNYEYAVGYWIKHGMSEKVMGIPYGHSEFLA  
SAETVGAAPSAGGAAGPTRESSGFLAYRIGFICGAKITRIODQOVPAYVGNOMV  
GYDVKSMERKVOFLKLNLGAMISIDMDPTGKSCNGPIPLVQAKRSLGS"  
77..106  
/gene="HUMTCHIT"  
/number=2  
107..309  
/gene="HUMTCHIT"  
/number=3  
115..1206  
/gene="HUMTCHIT"  
/note="first alternatively spliced form, see Genbank Accession Number U58515 for the second alternatively spliced form"  
/product="chitinase"  
310..365  
/gene="HUMTCHIT"  
/number=4  
366..516  
/gene="HUMTCHIT"  
/number=5  
517..643  
/gene="HUMTCHIT"  
/number=6  
644..771  
/gene="HUMTCHIT"  
/number=7  
772..954  
/gene="HUMTCHIT"  
/number=8  
955..1071  
/gene="HUMTCHIT"  
/number=9  
1072..1433  
/gene="HUMTCHIT"  
/number=10  
1210..1433  
/gene="HUMTCHIT"  
1407..1412  
/gene="HUMTCHIT"  
1434  
/gene="HUMTCHIT"  
/note="32 A nucleotides"  
BASE COUNT 375 a 360 c 342 g 357 t

ORIGIN  
Query Match 9.7%, Score 159; DB 74; Length 1434;  
Best Local Similarity 63.0%; Pred. No. 5,02e-128;  
Matches 711; Conservative 0; Mismatches 390; Indels 27; Gaps 10;  
Db 69 ctggcagggtagtggtctgtctgtctccaggaggatctctcaaacgtgttg 128  
19 ttt 78  
129 ctacttaccacactggtcccgagcgcgaggaaccaggaattccacccctgaatat 188  
79 CTACTTACCACACTGGGCCCGGACGACAGAGGAGGAGGAGGAGGAGGAGGAGG 138  
189 tgaaccttccatgctctcatctcatctcatctcatctcatctcatctcatct 248  
139 GGACCCGACGCTTGACACCCACGACCTCATGAGGCTTGCTGCTGCTGCTGCTG 198  
249 tatcataaggaacagagtgatgatgatgatgatgatgatgatgatgatgatgat 308  
199 GAGCACCAGCTGAGTGAATGACGAGACTCTTACACGAGATTCAATGAGCTGA 258  
309 gaatcccaactgaaattctctctgtccattgagaggttacctgtgttccaaagg 368  
259 GAATCCCAACTGAAGACCTGTTAGCCATCGAGGCTGGAATTTGGCACTGAGAG 318  
369 caacctatggtgattctctcatcatcatcatcatcatcatcatcatcatcat 428  
319 CACAGATATGATGACCCAGGCGCAACACGCTGACACTTGTCACTGGGCATCAG 378  
429 tctgaggaacataacttctatgactcgtgatgatgatgatgatgatgatgatgat 485  
379 TCTGGCAAAATACAGCTTTAGCGCTTGAACCTTGATGGAGATCCAGAGAAC 438  
486 -agaa-----a-acactat---ttcactgtgctgattcatatgattagcagaagc 533  
439 GAGCCCTGCCGTAGACAGAGAGCGCTTACACACCTGTGTACGAGACTTGGCCA 498  
534 tcagaaggacttcaaaaatccaaagaagaagcttctctgtgctcgggcgtatctgc 593  
499 CCACAGAGAGAGCCAGACCTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558  
594 agggaggcaaatgattgataacagctatcgaagtgaagaactggcaaaatctgatt 653  
559 TGGGCAACCTATGTGATCTGTGATCGAGTGAATAATGCCCGACGACCTGAG 618  
654 catcaacctctgtccttctacttccatcgaagtcttggaagaagcccttcaactgc 713  
619 TGTCAACCTTATGCGCTACGACTTCCATGCTCTTGGGAGAGG-----TCAC 672  
714 caacagcccttgcagcaagggtgagcagagcagagcagaagctcctactacagtgc 773  
673 TAACAGCCCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732  
774 atagctgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 833  
733 TGCTGCTGTCAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792  
834 cccacacatgaggcactcttcaacactggtc-tg--cagaaaccccggtggggccc 890  
793 GCCTACTAGGAGAGCTCTTACACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 852  
891 tgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgcct 950  
853 AGCACAGAGGCTGTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 912  
951 tggatattgcctgcttgccttgccttgccttgccttgccttgccttgccttgcct 1010  
913 TGAAGTCTGC--TCCGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 969  
1011 ctacgacagtaagggagagcaggtggtggtggtggtggtggtggtggtggtggtg 1070  
970 CTACATCTTCGGGAGACACAGTGGGTGGCTTTGATGATGTGAGAGCTTCAAAAC 1029



LOCUS	6	SSGP38KD	1733 bp	RNA	MAM	28-AUG-1996
DEFINITION		S.scrofa	38kDa	heparin-binding	glycoprotein.	
ACCESSION		Z47803				
NID		9634097				
KEYWORDS		gp38k; heparin-binding				
SOURCE		plg.				
ORGANISM		Sus scrofa				
REFERENCE		Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Suiiformes; Suidae; Sus.				
AUTHORS		1 (bases 1 to 1733)				
TITLE		Millis,A.J.				
JOURNAL		Submitted				
REFERENCE		Submitted (19-JAN-1995)				
AUTHORS		Biological Sciences, 1400 Washington Avenue, Albany, NY, USA, 12222				
TITLE		Shackelton,L.M., Mann,D.M. and Millis,A.J.				
JOURNAL		Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in				
MEMLINE		differentiating vascular smooth muscle cells as a member of a group				
FEATURES		J. Biol. Chem. 270 (22), 13076-13083 (1995)				
source		9528589				
		Location/Qualifiers				
		1..1733				
		/organism="Sus scrofa"				
		/clone="pBS38k"				
		/dev_stage="adult"				
		/tissue_type="smooth muscle"				

	CDS	/clone_1lb-"1ambda ZAP II"
		67..1218
		/codon_start=-1
		/product="38 kDa heparin-binding glycoprotein"
		/db_xref=" ID:J634098"
		/translation="MGLRVAQIGFVALVLLGSCAAKYLCYTTSSQVRREGSGCF AIDPDLCTHIIYSFANIISNNEIDTLEMNVDLTLDTLNLKNNPIKLTVLGGWNN SQRFSLASNTSFRRTFKISVPFLTFDFDGLDIIDLMSLRBRDLTKLVKKKK FVEALDPGETRLLSGAVSAKAVIDRGADLOIASHDFISLYTFDFGARQRTT HSPLPRGOGASSDRFSNMADVANSYLRLGIAPANKHWMLIPFGESFTLASKTDV PASGPGRGFETFEKELIAYEICDQLQATRYRRPGOVYPATNGNWCGIDDBE KNAKYIKRSQRLGAWMTLDDDERGNRCGONLPRLTSALKDYLAAR"
B	BASE COUNT	408 a 507 c 463 g 355 t
O	ORIGIN	
Q	Query Match	9.5% Score 156; DB 48; Length 1733;
B	Best Local Similarity	62.5%; Pred. No. 6,59e-125;
M	Matches	688; Conservative 0; Mismatches 388; Indels 24; Gaps 9;
D	Db	gtggtcagagaaggcttttgttgcccttggctctgtgccacagaccygtgtgcctcaaacgtc 138
O	14	GTGGCTGGGAGAGGTTTCATTGCTCTCTCTATATCCCATGGGGCTCTGTCGAACCACTG 73
D	Db	gttgctactatlaacaaagtctgttcagtlaccgggaagtgatgtggaagctgtcccagat 198
O	74	GTCCTACTACTCACCAACTGGGCCCATACAGACAGGGGGAGGCTGCCTCTCTGCCAAC 133
D	Db	gccatcgcccccttcctctctgaccaccaatacatctaacagctttgccacaataagaacaat 258
O	134	GACTTGAGACCACCTTTTGACCCACCTCATATACCCCTTCCCTGGCATAAGCAACCAAC 193
D	Db	gagattgcacacctggagtggaatgatgttaagctctcatagacaactgaaacacccaag 318
O	259	GAGCTGAGACACACTGAGTGGAAATGAGAAGACTCTCTACAGAGAGTTCAATGGCCCTGAAG 253
D	Db	aacaggaaccccccaactgaaagaccctctctgtctgtttggaatgaaacttggcttcaa 378
O	254	AAGATGATCCCAAGCTGAAGACACCTGTATAGCCATCGAGGCTGGAATTTGGGCACTCAG 313
D	Db	agattttccaataaatagctcccaaacactcagaagtctgcgaagacttcatcaagtctgtgcca 438
O	314	AAGTTCACAGATATGTAAGCCACGCGCACMACCGTAGACCTTTGCAACTGGCCATC 373
D	Db	ccattccctgagaccatggtgttatgtagactggaacctagccctggatctccctgg~9c 496
O	374	AGGTTTCTGCGCCAATACAGCTTTGACGGCCTTGACCTTGACTGGGAGTACCCAGGAAGC 433
D	Db	-----agaaaagcggcatctcacacactagctcaagaaagatgaaagct 543
O	434	CAGGGGACCCCTGGCGTAAGACAAGACGCGCTTCCACACCCCTGGTACAGACTTGGCCAT 493
D	Db	gagtttgtaaaggaagccctg~ect--ggaacagaagcgtctctgtctcagcggagcgtg 600
O	494	GCTTCCACAGCGGAAGACCCAGACCTCTCAGGAAGGAAGAACCTTCTTGATGATCGACGGTT 553
D	Db	tctgcagagggaagrgcgcacattgaaaggggtatgacctgcgccagatatctcaaacctcg 660
O	554	CCAGCTGGGAGACCTATGTGGATGCTGTGATACGAGGTGAACAANAATCGCCCAAGAACCTG 613
D	Db	gaattcaacaacetttyagactatgacttcaacggtctcggcgcccaaacattcaggacac 720
O	614	GATTTTGTCAACCTTATGGCTACAGATTCTTCATAGCGCTTGGGAGAAAGTCAAGGGACAT 673
D	Db	cacagtcceccctgttcsgaagccagggggatgtaagttccgacagatcagcaatgacag 780
O	674	AACAGCCCCCTCTCAAGAGGCAAGAGAAGATGGTGCACACACCGACCTCAACGCTGAT 733
D	Db	tatgctgtgagactgtgtgctggaagcttggggcccccagccaataagctgtgtagtggatc 840
O	734	GCTGCTGTGCAACAGTGGCTGCAAGAAAGGAGACCCCTGCACGCAAGCTGATCCTTGGCATG 793
D	841	cccaactttgggaagagcttcaacttggcctcttcca~aagacag~ac~gtggagagcccg 897

[illegible]

	RESULT	8	
ID	SSGP38KD	standard; RNA; MAM; 1733 BP.	
AC	Z478037		
N1	G634097		
DT	19-JAN-1995 (Rel. 42, Created)		
DT	28-AUG-1996 (Rel. 49, Last updated, Version 6)		
DE	S.scrofa 38kDa heparin-binding glycoprotein.		
KW	gp38k; heparin-binding glycoprotein.		
OS	Sus scrofa (domestic pig)		
OC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
OC	Theria; Eutheria; Artiodactyla; Sulfurformes; Suidae.		
RN	[2]		
RP	Mills A.J.;		
RA	Mills A.J.;		
RT	; Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.		
RL	Mills A. J., University at Albany, Biological Sciences, 1400		
RL	Washington Avenue, Albany, NY, USA, 12222		
RL	[3]		
RP	1-1733		
RX	MEDLINE; 95286589.		
RA	Shackelton L.M., Mann D.M., Mills A.J.;		
RT	"Identification of a 38-kDa heparin-binding glycoprotein (gp38k)		
RT	"in differentiating vascular smooth muscle cells as a member of a		
RL	group of proteins associated with tissue remodeling";		
RL	J. Biol. Chem. 270:13076-13083(1995).		
FH	Location/Qualifiers		
FT			
FT	source		
FT	1..1733		
FT	/organism="Sus scrofa"		
FT	/clone="pBS38k"		
FT	/dev_stage="adult"		
FT	/issue_type="smooth muscle"		
FT	/clone_lib="lambda Zap II"		
FT	67..1218		
FT	/product="38 kDa heparin-binding glycoprotein"		
FT	/db_xref="PID:g634098"		
FT	/translation="MGLRVAQTGFVALVLQSCAAYKLVCYYTSMQYREGDSGCFPD		
A			
FT	IDPFLCTHIISFANISNNIEDTLLEMDVLTLYDILTNLTKNNRPLKTLLSVGGNPFSS		
FT	O		
FT	R	RFSFIASNTQSRRFFIKSVPEFLTTHGEFDGIDLAMISPGRDRKHLLTYLKEMKAEEVT	
FT	L	EALPTGERLLSLGSASAGKVAIDRGYDIAQIQHLIDELISLTYDFHGAMROTTHHSPP	
FT	P	FRGGDASSDRFSNADYAVSYVLRGLGAPANKLVWGIPTFGKSFTLASSKTIDYGAPASGS	
FT	Y	GIPRFPRFEKGIILAYEELCDPLQGATVRRPLGGQVPATKGONMWGYIDDQESYKNKKAR	
FT	Sequence 1733 BP; 408 A; 507 C; 463 G; 355 T; 0 other:		
FT	SQ	LKSRQLAGAWMYTDLDDPFGNFCGOMLRPELTSAINDVLAAG	
	Query Match	9.5%; Score 156; DB 15; Length 1733;	
	Best Local Similarity	62.5%; Pred Nisat 6.59e-125;	
	Matches	688; Conservative 0; Mismatches 388; Indels 24; Gaps 9;	
Dd	79 gtgctcagacagggttttgccctggctgctgcctccagagctgtgctgcttaccagctg	138	
Oy	14 gttgccctggcgaggttcattgatgctctgctgtagatgccatggagggctgctgcamaacg	73	
Dd	139 gttgctctatacacagctggtctcaagtaccggggaggtgtagtgggaagctgtctccaat	198	
Oy	74 gttcctacttcttaccacctggggccagtaacaacaggggagggcttccgcttccgccaag	133	
Dd	199 gccacacgcccccttcctgcaccacacatcatcatacaagctgttgcacaataagaacaat	258	
Oy	134 gacttggaaccccagccttgtagcaccacacatcatcatgacgcttggcggatgacaaacac	193	
Dd	259 gagattgacaccttgtagtgaatgtagtgcgccttatgacacactgaacacactcaag	318	
Oy	194 cagctgacacacacagagtgagatgacgagacttcttaccagagagttatagcccttgag	253	

LOCUS	9	1616 bp	RNA	ROD	28-NOV-1995
DEFINITION	M. musculus mRNA for BRP39 protein.				
ACCESSION	X93035				
NID	g1085065				
KEYWORDS	brp39 gene; BRP39 protein.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorphia; Muridae; Murinae; Mus.				

REFERENCE	1 (bases 1 to 1616)
AUTHORS	Morrison, B.W. and Leder, P.
TITLE	neu and ras initiate murine mammary tumors that share genetic markers genetically absent in c-myc and int-2-initiated tumors
JOURNAL	Oncogene 9 (12), 3417-3426 (1994)
MEDLINE	95060797
REFERENCE	2 (bases 1 to 1616)
AUTHORS	Morrison, B.W.
TITLE	Direct Submission
JOURNAL	Submitted (14-NOV-1995) B.W. Morrison, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES	Location/Qualifiers

```

/organism="Mus musculus"
/strain="FVB/N"
/sex="male"
/age="12"
/tissue_type="adult"
/tissue_type="breast"
/cell_line="SKF murine breast tumor"
/clone_id="lambda Zap II"
66..128
misc_feature

```

2

/note="homology to glycosyl hydrolases"  
 /codon\_start=1  
 /product="HRp39 protein"  
 /db\_xref="pid:e211795"  
 /db\_xref="pid:g1085066"  
 /translation="MGMRALATGFRAVIMLLQSCAAYKLYCVFTSMQYREGCVGFLEP  
 AIQPLETHIIYSPANISDDMLSTWENDEDSNDYKLKLTNTNMLTLLSYGKWKKE  
 GEKFESEIATSTERTALFASVAPFLRSYGGDILAIYRLDRKQVFSTLLKEINLA  
 EFTKEVQSGREKLLISALISAGVAIDYGIQAIQALHLDPIINLTMYTFHFWMOITG  
 HHSPLFGQKDRTREDRYSNVAYOVMYRLGAQASKLIMGIPTGKSPITLASNNQAG  
 AHSIGGLPGRFTREAGTLAYEYCDPLKQGEVHYRLSMEKIPFATKGNQWYGEHKES  
 VKNVYGFUKERKLGAAMWALDLDFOCTQPKFEPLTNAIKDALA"

Query Match	8.3%;	Score 135;	DB 81;	Length 1616;
Best Local Similarity	62.5%;	Pred. No. 3.23e-103;		
Matches	674;	Conservative	0;	Mismatches 377;
			Indels	27;
			Gaps	13;

Db 88 cagagcttgcgagctccgtgctgtgtcccaagctctctgtgtaacaagtgtgtcgaact 147  
 Qy 24 CAGGTTTATGTGTCTCTGCTATGATTCCTCAATGAGGGCTTGCTGCGAAAAGTGTCTGCTACT 83  
 Db 148 tcaccagctgtgtcccaagctaccgagggaagcggttggaaagctctcttaccagagccatccaac 207  
 Qy 84 TCACCAACTGGGGCCOAGTACAGACAGAGGGGAGGCGTCCGCTTCCTCTCCCAAGAGACTTGGACC 143  
 Db 208 ctcttcctgttgcaaccaacaatcatctacagctttgccaacatgaagcagcaacaattgcta 267  
 Qy 144 CGAGGCTTTGGACCCACCACCTCATCTACGCGCTTCGCTGGCATGACCAACACC--CA-GCTGA 200  
 Db 268 gcaacttggagatlgatbtagcagagctgcgaactatgtcaagctcgtgaataaactaagaccagaa 327  
 Qy 201 GCACCACTGAATGTGATACACAGACTCTTACCGAGGATTCAAAGGCGCTCAAGAGAGATGA 260  
 Db 328 aacaccaacttgaaaacccctccctgcgtctgttggagggtggaaattggcgaaaaaagatttc 387  
 Qy 261 ATCCCAACCTGAAGACCCCTGTTAGCCATCCGGAGGCTGGAAATTCGGACATCAGAAAGTTCA 320  
 Db 388 ccgaagatgtccttcaacaacactagaagacgcaactgtgttcctgcgttcggttagccccgfttc 447  
 Qy 321 CAGATATGTATGCCACGGGCCAACAACGCTAGACCTTGTCAACTCCGGCATCAGGTTTC 360  
 Db 448 tgcgttctctatlgacttltgatgtggctggaatctgcgctgtgcttaccct-----c----- 495  
 Qy 381 TGGCGAATATACAGCTTTTACGCGCTTGACCTTGAATGGAGATACCCAGGAAGCAAGGGA 440  
 Db 496 gct-taa-g-agaacagcagatattcttccacccttgatcaagaagaaactaatgtcggaattca 552

Oy	441	GCCTGCGGTAGACAAGAGAGCGCTTACAAACCTGGTAGACAGACTTGGCCAAATGCTTCC	500
Db	553	caaaaggagttccag--c-caaggcagagagaactcctgctcagcgcaagctttgtcagcag	609
Oy	501	AGCAGAGAGGCCCAACCTCCAGGGAGGAAAGAACGCCCTTCTTCTTGAGTGGACGGTTCACAGTG	560
Db	610	gaaaaggagccattgacacctggcctatpacatgcgccagatagcccaacaccttgatttta	669
Oy	561	GGCAGACCTATGTGGATGCTGGATACGAGGTGGACAAATCGCCCGAAMCTTGATTTTG	620
Db	670	tcaattcattacactacgatttccatgagagctctgycgccaatacacaagccacacaagcc	729
Oy	621	TCAACCTTATGGCTTACGACTTCATCGCTTGTTGGAGAGAGTGCACGGGACATACAGCC	660
Db	730	ccctcttccaaagccagaaagaaactagtggttttacagatacagaatlytaactatgpcg	789
Oy	681	CCCTCTACAAAGGGCAAGAGAGAGTGGTGACAGACGCCCTCAACGTGATGCTGTG	740
Db	790	tgcagttacatgatcgcctctggagagcccaagccagcaagctcactgatyggatcccaact	849
Oy	741	TGCACAGTGGCTCAGAAAGGGAGCCCTTCACAGCAAGCTGATCTTTGGATGCTACCT	800
Db	850	ttgggaagagagttacactctgcatcttc-tyaaa-atcagc-tyggagatcccaatcag	906
Oy	801	ACGGACGCTCTTACACTGGCGCTCCCTCATACAGACACCAAGTGGGGCCCAACGCACAG	860
Db	907	gggaagagattaccagagccggtttccacaagaagagcaggaagccctgagccctactaagatat	966
Oy	861	GGTCGGGACCTCCAGGCCCTTCCACAAAGGAGGAGTGTGGCTACTATGAAAGTCT	920
Db	967	ggcacttccctaaaaggactgtaagatcatctgactcttcaagagaaggttcctctgcta	1022
Oy	921	GCTCTGTG--TAAGGGGGCCACCAACAGAGAAATCCAGGATCAGAGGTGCTCCCTACATCT	977
Db	1027	ccaaagggcaaccagttgggttggtgagtatagacaagaagagatgycaaaagaagttgggt	108
Oy	978	TCCGGGACACCACTGGGTGGGCTTTATGATGTGGAGACTTCAAAACCAAGTCAAGCT	1033
Db	1087	tctctgaaggagaagaagctgtgcagagagcatalgtgtgagcactgatttgaatgattt	1144
Oy	1038	ATCTGAAGCAAAAGGAACTGGCGGGGCCATGGTGTCTGGCAGCTTGAGTGAAGCTT	1095
RESULT	10		
ID	MMD757	standard; RNA; ROD; 1506 BP.	
AC	D87757;		
NI	g1545818		
DT	15-SEP-1996 (Rel. 49, Created)		
DT	15-SEP-1996 (Rel. 49, Last updated, Version 1)		
DE	House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L precursor, complete cds.		
KW	ECF-L precursor.		
OS	Mus musculus (mouse)		
OC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
CC	Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.		
RN	[1]		
RP	1-1506		
RA	Owashi M.;		
RT	;		
RL	Submitted (06-SEP-1996) to the EMBL/GenBank/DBJ databases.		
RL	Macro Owashi, The university of Tokushima, Faculty of Integrated		
RL	Arts and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770,		
RL	Japan (E-mail:owashi@etas.tokushima-u.ac.jp, Tel:0886-56-7261,		
RL	Fax:0886-56-7298)		
RN	[2]		
RP	1-1506		
RA	"Owashi M., Arita H., Taoka Y., Hayai N.;		
RT	"Molecular cloning and characterization of T lymphocyte-derived		
RT	eosinophil chemotactic factor (ECF-L)",		
RL	unpublished.		
TH	Key	Location/Qualifiers	
TH	source	1..1506	



FT		/organism="Mus musculus"
FT		/cell_type="bone marrow cells"
FT		/tissue_type="bone marrow"
FT	CDS	1..1197
FT		/gene="meef-L"
FT		/codon_start=1
FT		/product="ECF-L precursor"
FT		/db_xref="PID:g1545819"
FT		/translation="MAKLIVLTGLAILNLVOLGSSYOLMCYYTSMANDRPIEGSFKPGP
FT	N	IDPLCHLIIFAFGMQNNELTYTHEDLRVYEAHLNGAKDNTSLKTLILAIGKKFPGR
FT	A	SFSAMVSTPNQRQIFIOSVIRFLROYFNFDGLNDMQYPSGSFPPKKHLSVLVKEM
FT	R	KAFEESEVERKDIPRLILNSTAGIIDIYIKSGKIPEISQSUDYIQVMTYDLHDHPDKDY
FT	T	GENSPLYKSPYDIKRSADLNDSDSIYSWKDHGAASEKLIVEFPAYGHFTILSDSKTG
FT	I	GAPITISGPPEPKYIDESGLAIYEVCFFLNEGATEWMDAPEVPYAYQENWGYNV
FT	R	SFKIAQMWLKNNNLGAVWMVRLLDMDDSGSFCHORHFPLSTLKGLDINHSAOSCKPY
SO	Sequence	1506 BP; 422 A; 337 C; 335 G; 412 T; 0 other.
	Query Match	7.5%; Score 123; DB 20; Length 1506;
	Best Local Similarity	57.2%; Pred. No. 5,92e-91;
	Matches	610; Conservative 0; Mismatches 451; Indels 6; Gaps 4;
D	70	ctgaatgctactatacaagttggcctaaggccaatatgaaggagtgttaacct 129
O	71	CTGGTCCTACTTACCAACTGGGCCCATGATACAGCAGGGGAGGCTCGCTCCGCC 130
D	130	gataattgcccccgcctgctgctacaccatgactatgaccttccttgctggaatgcagaat 189
O	131	AAGSACTTGGACCACCGAGCTTTGCACCCACCTCACTTAGCCTTGCTGGCATGACCAAC 190
D	190	aatgatcatcttacacacatgaaagaacttcgtgactatgaaecatltgaatgctg 249
O	191	CACCGACTGAGCACCATGATGTGAATGACGAGACTCTCTACAGAGATTCACTGGCCTG 250
D	250	aaagccaagaacaactggaactaaaactctccctggccattggaagatlgaaatttggacct 309
O	251	AAGAAGATGAATCCCAACACTGAAAGACCTGTTAGCCATCGGAGGCTGGAAATTTCGGCACT 310
D	310	gcctcgctcagtgcatgctgctcactcctccgaagccctcaagatatcatcaafgcagt 369
O	311	CAGAACTTACAGARATGTGTAGCCACGGCCAAACAACCTGTGTAACCTGGCCC 370
D	370	atcagaatccttcgtcaaalataactltgatggtccctcaaaccttgaacttgcagttacccctg 429
O	371	ATCAGATTTCGGCAATAATACAGCTTTGACGGCTTGACCTTGACTGGGAGATACCCAGGA 430
D	430	tctcgaaagaagccctcctcaaggacaaacatctcttcagttgctcgttgaaggaaatgcgt 489
O	431	AGCCAGGGGAGCCCTCCGCTAGACAAAGAGACCTTCAACAACCTGTGTACAGACTTGGCC 490
D	490	aaagcttttgaagaagatctgtggaagaagacatccaaggctgcctactcactccaca 549
O	491	AATGCTTTCACAGAAAGAACCCAGACCTTCAGGGAAGAAAGCGCTTCTTGTGATGCTACAG 550
D	550	ggagagaagaatcatgtagcgtlaatacaagtcgtgtatacaagaatccctgaactgctcagct 609
O	551	GTTCAAGCTGGGCAACCATATGTGTACTGTGATACGAGTGTGACAAAATGCCCCAAGAC 610
D	610	cctgaactatattcgaagtcatagacatatgattcctcatgactccaattaagatggtcaactgga 669
O	611	CTGGATTTTGTCAACCTTATGCTTACGACTTCCATGCTCTTGGGAAAGAGTCCAGGGA 670
D	670	gaaaatagtcctccctcataaactcctcatatgcatcttggaagaagagcgtcatcaattg 729
O	671	CATTAACACCCCTCTTACAGAGGCAAGAGAGATGTGGACGACGACCAAGCCTCAACGTG 730
D	730	gattcaacattctctactggaaggacatgagcagcttcctgagaagctcatatttgga 789

QY	731	GATGCTGCTGTCAACAGAGTGGCTGCAGAGAGGGAGCCCTCGCAGCAAGCTGATCCTTGGC	790
Db	790	ttccagcatataggccataccttacccttccagtgagtgaccccttcaagactggaattgtgac	849
QY	791	ATGCTTACCTACGAGCAGCCTCTTCACTAGCTGGCTCTCTATCAGACACCAAGATGGGGGCC	850
Db	850	cctacaattagtagctgagccaccaggaagtacacagatgaactgaactggaactcctggttac	909
QY	851	CCAGCCACAGGGGTCTGGCAGCTCCAGGCCCTTCACCAAGGAAGGAGATGCTGGCTTAC	910
Db	910	tatgaggttctgacatctcctgaatgaaggaagcactgagtggtctgtagtgcctccagaa	969
QY	911	TATGAAGACTG--C--TCTGTGAAGG-GG-GCCACCAACAGAGATCCAGATCAGAG	964
Db	970	gtacccttgctacccaaggtgaatgagtggtgtgttgaacaaagtcaagagcttcaag	1029
QY	965	GTGCTTCATATCTTCCGGACACACCAAGTGGGGCTTTGATGATGTGAGAGCTTCAAA	1024
Db	1030	ttgaagagctcagtggtgctcaagagacaacaaattagaaggtgcctgtgtcgtgcctgagc	1089
QY	1025	ACCAAGGTACAGTATCTGAAGAGAAAGGAGACTGGGGGCGCATGTCTGGCAGCATGGAC	1084
Db	1090	atgagatcctcaatggtgtctcttctgtcaccaagagaatcttcctct	1136
QY	1085	TTAGATGACTTTGGCGGCTTCTCTCTGCAACACGAGGCGCATACCCCTC	1131
RESULT	11		
LOCUS	D87757	1506 bp	13-SEP-1996
DEFINITION	House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L precursor, complete cds.		
ACCESSION	D87757		
NID	g1545818		
KEYWORDS	ECF-L precursor.		
SOURCE	Mus musculus bone marrow cells cDNA to mRNA.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 1506)		
AUTHORS	Owashi, M.		
JOURNAL	Direct Submission		
TITLE	Submitted (06-SEP-1996) to the DDBJ/EMBL/GenBank databases. Makoto Owashi, The University of Tokushima, Faculty of Integrated Arts and Sciences, 1-1 Minami-Johsanjima, Tokushima, Tokushima 770, Japan (E-mail: owashi@ias.tokushima-u.ac.jp, Tel: 0886-56-7261, Fax: 0886-56-7298)		
REFERENCE	2 (bases 1 to 1506)		
AUTHORS	Owashi, M., Arita, H., Taoka, Y. and Hayai, N.		
TITLE	Molecular cloning and characterization of T lymphocyte-derived eosinophil chemotactic factor (ECF-L)		
JOURNAL	Unpublished (1996)		
FEATURES	Location/Qualifiers		
SOURCE	1..1506		
	/organism="Mus musculus"		
	/cell_type="bone marrow cells"		
	/tissue_type="bone marrow"		
	1..1197		
	/gene="mECF-L"		
	/codon_start=1		
	/product="ECF-L precursor"		
	/db_xref="PID:g1545818"		
	/translation="MAKULILVTGAILLNLVQGSYOLMCTYTSWAKDRPIEGSFKPG		
	NIDSLCTSLIYAFAGQNNETITTHODLNDYALNLGQDKNLELTLLAIGKRGK		
	PASISAVSTPQKQIRFQSVIRFLRYNFGGLMDQVPSRSPKDKHLEVLVK		
	EMRAAESESEYKNDIPRLILSTGAGLIDVKSXKIPELSSQIDVQVNTYDLHPR		
	DYGVGNSPLEKSPYDYGKADLVNDSIISYMRHGAASEKLIVFPAYGHTFLSDP		
	SKTGIDQISTSGPCKYTDSEGLIAYVCTPLNEGATEVMDPAQEVPAAYQGNEMV		
	GYDVRSFKLAKQWLKDNNGAGAYVWMLDMDPFGSFCHQHFPLTSLKGLNISHA		
	SCROPY"		
BASE COUNT	422 a	337 c	335 g 412 t
ORIGIN			





ACCESSION	partial cds.
NID	017988
KEYWORDS	9602293
SOURCE	sheep.
ORGANISM	Ovis aries
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprine; Ovis.
AUTHORS	1 (bases 1 to 1596)
TITLE	Marshall,J.T., Brownlee,A.G. and Nancarrow,C.D.
JOURNAL	Cloning and sequencing of a cDNA encoding ovine oestrous glycoprotein
REFERENCE	Reprod. Fert. Dev. 8 (2), 305-310 (1996)
AUTHORS	2 (bases 1 to 1596)
TITLE	Marshall,J.T.A.
JOURNAL	Direct Submission
REFERENCE	Submitted (02-DEC-1994) James T. A. Marshall, Division of Animal Production, Commonwealth Scientific and Industrial Research Organization, Locked Bag 1, Delivery Centre, Blacktown, New South Wales, 2148, Australia
FEATURES	Location/Qualifiers
source	1..1596
	/organism="Ovis aries"
	/strain="merino"
	/sex="female"
	/cell_type="epithelial cell"
	/tissue_type="ampulla oviduct"
	/dev_stage="oestrus"
	<1..1596
	/codon_start=1
	/product="92kd oestrus-associated oviduct glycoprotein"
	/db_xref="PID:9602294"
	/translation="LLVYLKHHDAHKLVCYFTNMAFSRPGSAILPRDIDPLCTH
	LVEFARNNNQIYKDPIDELKILYPEENKKEKNGKATLISGVNGFSGRFTKM
	TSFNRRENVSVIALRTGHEFDGLDFEYLPGLRABDMFVLEIELLQAFKK
	EAQITLMPRLILSAVSGDPIYIKADARLQGLDPLISLSDIHSWEKTYGHS
	PLFELPDPKSAYAMSTYRQLGVPPELILGLPTIGTHLRLASQDLGAVGPP
	SPGVTYKQAGFLAYEVCSFVQRAKKRINQYVYPAKGEWGYDAISFGKAFEE
	IKRHFGGAWMTLDDLFRCNPGCGTFEPLVHTLNNLLVNDSESTPSPFWSTAW
	NSISIGEMPMPTDLITGLGILPLGGEAVATEHRKSAWTTPRGETATPTPLSL
	SGRTAAPEKTESPERKPLRVSGLHVAVPGIAGVPHLQIGKVPDGRKAGVPE
	VTMSGKWYTPDGRATLERRL"
mRNA	<1..>1596
	/product="92kd oestrus-associated oviduct glycoprotein"
	<1..36
	/note="Inferred from bovine sequence: GenBank Accession Number D16639"
	/codon_start=1
	1..19
	/note="derived from PCR primer based on bovine CDNA sequence: GenBank Accession Number D16639"
	/note="derived from PCR primer based on bovine CDNA sequence: GenBank Accession Number D16639"
BASE COUNT	352 a 440 c 430 g 374 t
ORIGIN	
Query Match	5.7%; Score 93; DB 47; Length 1596;
Best Local Similarity	58.7%; Pred. No. 6,99e-61;
Matches	364; Conservative 0; Mismatches 255; Indels 3; Gaps 2;
Db	40 aagctggtgtgtattattaccacaactggtgcatcagtcggccggtctgctcgatctcg 99
QY	68 AAACGAGCTGCTACTTCACTCAACATGCGGCCAGTACAGACAGGAGGAGGCTCGCTCC 127
Db	100 ccccgagatctggacccttcttcgtgacacccacacctggtatttgctcttgctcgatgaac 159
QY	128 CCCAAGGACTTGAGACCCACGACCTTGACACCCACCTCATCTAGCGCTTCCTGGCATGAC 187
Db	160 aacatcgatgtttccccaagatccccggatgtagaagaatccttaccacagatitcaac 219
QY	188 AACCAACGAGCTGAGACACCACTGATGAGTGAAGAGAC--TC-TCTACCAAGAGATTCAAT 244

Db	220	aagctcaaggagaaggaacaaagggcgtaaacacactgctgcgttgaggggatgaactc	279
Oy	245	GGCCTGAAAGAAGATGAATCCCAACGTBAABACCCTGTATTACCATTGCGAGGCTGGAAATTTC	304
Db	280	ggcacctcagagtctaccaaagatgctgccagcttctccaacccggaaaagtgttcnat	339
Oy	305	GGCACCTCAGAAGTATCACAGATATGTGTAGCCAGGCCAACCAACCGTCAGACTTTTGTAAC	364
Db	340	tcaagtatgcacctccctcaggaagaacatagsgttcgtgatgattgttgtgaacctcttcttgtag	399
Oy	365	TGGGCCATCAGAGTTCTGCGCCAATAATACAAGCTTTGACGGCCCTTGACCTTGACTGGAGTAC	424
Db	400	ccttgatacagaagacaagccccgcgcgtgacgcctctgagaccttctctctaacttgagag	459
Oy	425	CCAGAAGACCGAGSGGACCCCTGCCGTTAGAACAGAGAGCGCTTCAACACCTCGTATACAGAC	484
Db	460	ctctctcagagcctccaagaatagagcccacataccatctgcgcccaagagctgctgctc	519
Oy	485	TTGGCCATATGCTTCCAGAGGAAGCCCCACCTCAAGGAGGAAGAACCCCTTCTTGAGT	544
Db	520	gctcgcctctctgaggccccccacgcatatccagaagcatatgatgacgccttcgggc	579
Oy	545	GCACCGGTTCACGTTGGCGAGACCTATGTGATGCTGTGATACAGAGGTGACAAAATCGCC	604
Db	580	agactcctgattcatcaccagcgtctgtcttaagacttacacggaagcgtgaaaaagtc	639
Oy	605	CAGAACTCGATTTTGTCAACOTTTATGGCTACGACATTCATGCTTGTGGAGGAAGTCTC	664
Db	640	acagagacacaatagccctct	659
Oy	665	ACGGACATATAACAGCCCCCT	684
RESULT	15		
LOCUS	BABED0SG	2228 bp	mRNA
DEFINITION	Papio hamadryas anubis estradiol-dependent oviduct-specific glycoprotein mRNA, partial cds.		
ACCESSION	M59903		
NID	g1256950		
KEYWORDS	estradiol-dependent oviduct-specific glycoprotein.		
SOURCE	Papio hamadryas anubis cDNA to mRNA.		
ORGANISM	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.		
REFERENCE	1 (bases 1 to 2228)		
AUTHORS	Donnelly,K.M., Fazleabas,A.T., Verhage,H.G., Mavrogianis,P.A. and Jaffe,R.C.		
TITLE	Cloning of a recombinant complementary DNA to a baboon (Papio anubis) estradiol-dependent oviduct-specific glycoprotein		
JOURNAL MEDLINE REFERENCE	MOL. Endocrinol. 5 (3), 356-364 (1991)		
AUTHORS	2 (bases 1 to 2228)		
TITLE	Jaffe,R.C., Atlas,E.B., O'Day-Bowman,M.B., Donnelly,K.M., Mavrogianis,P.A. and Verhage,H.G.		
JOURNAL	Regional Distribution and Hormonal Control of Estrogen-Dependent Oviduct Specific Glycoprotein mRNA in the Baboon (Papio anubis)		
REFERENCE	Biol. Reprod. (1996) In press		
AUTHORS	3 (bases 1 to 2228)		
JOURNAL	Jaffe,R.C.		
REFERENCE	Direct Submission		
TITLE	Submitted (05-MAR-1991) Randal C. Jaffe, Department of Physiology and Biophysics, University of Illinois College of Medicine,		
JOURNAL	Chicago, IL 60612-7347, USA		
FEATURES	Location/Qualifiers		
SOURCE	1..2228	/organism="Papio hamadryas anubis"	
CDS	13..1884	/codon_start=1	
		/product="estradiol-dependent oviduct-specific glycoprotein"	
		/db_xref="PID:g1256951"	







QY 84 TCACCAACTGGGCCGATACAGACAGGGGGAGGCTCGCTTCTGCCCCAAGCACTTGAGCC 143  
Db 214 gcttcctgtgtacccatcatcatcagcttgcacatataagcaaacatcacatgcaca 273  
QY 144 CCAAGCCTTTCACCCACCTATCTACGCCCTTCCGTGGCATGACCAACACACAGCTAGCA 203  
Db 274 cctggagatggaatgatagtacgctctacagcaltgctcaaacacatcaaacagcaacc 333  
QY 204 CCACTGAGTGAATGACGAGACTCTCTACAGAGAGTTCATGAGCCCTGAGAGAGATGATC 263  
Db 334 ccaactggaagactctctctgtctgtcggaagatggaacttgggtctcaaatattcca 353  
QY 264 CCAAGCTGAAGACCTGTTGACCATCGAGCTGGAATTTCCGACATCAAGTTCACAG 323  
Db 394 agatagctctcaaacacagagatgcgcggaacttcatcaagtaagtaacgcacatttgc 453  
QY 324 ATATGTAAGCCAGGCCAACACCGTTCAGACTTTGCAACTGGCCATCAGGTTTCTGC 383  
Db 454 gcaacccatggtcttgatgagcgatgaccttgcgtctacacccctgga---c---ggaag-- 505  
QY 384 GCAAAATACAGCTTTGACGGGCTTGACCTTGACTGGAGATGCCAGAAAGCCAGGGGAGCC 443  
Db 506 -----agcaaacacacatttaccacccctaatcaaggaatgaaagccgaatttata 558  
QY 444 CTGCCGTAGACAAAGAGAGCGCTTCAACACCTGTACAGGACTTGCCCAATGCCCTTCAGC 503  
Db 559 aggaagcccgag--c-cagggaaaaaagcagctcctgcgtcagcgacgacgtgtctgcggga 615  
QY 504 AGGAAGCCCAAGACTCTCAGGAGAAAGAACGCTTCTCTGATGTGACGGCGTTCAGATGAGC 563  
Db 616 aggtacacattgacagcagatgacatgacatgacagatatacccaacacgtgatttca 675  
QY 564 AGACCTATGTGATGCTGATACGAGAGGTGAGCAAAATGCGCCAGAAACCTGATTTGTCA 623  
Db 676 gcatcaatgacctagatttcatagcgccgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 735  
QY 624 ACCTTATGGCTGACGATTCATGAGCTTCTTGGAAGAGTCCAGCGGACATACAGGCCCC 683  
Db 736 tgttcgaggtcaggaagatgacagatgcacagattcagcaaacacgtgatttgcgtg 795  
QY 684 TCTACAAAGAGGCAAGAAAGATGCTGACAGCAGCAGCTCAGCTGATGCTGCTTGC 743  
Db 796 ggtacatgttgaagcgttgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 855  
QY 744 AACAGTGTGTCAGAAAGAGGAGCCCTGCAAGAGTGTGATGCTGATGCTGATGCTGATG 803  
Db 856 gtaggaggtcactctgtgctcttc--tgagacgtg--gtgttcagcgcaatctcaagac 912  
QY 804 GAGGCTTCCTTCACACTGGCTCTCTCATACACACAGATGAGGGGCCCCAGCCACAGGCT 863  
Db 913 cgggaattcgaagccggtttacaaaggaagcaggaacctgacctactatgagatctgtg 972  
QY 864 CTGGCACTCCAGGCCCTTTCACCAAGAGAGAGGAGATGCTGAGCTTATGATGATGCTGCT 923  
Db 973 acttctctcgcgcgcgcacagatccatagaacctgcgcgcgcgcgcgcgcgcgcgcgcgc 1032  
QY 924 CCTGGAAG-G-GG-GCCACACAAACAGAGAAATCCAGATCAGAGAGTCCCTTACATCTTCC 980  
Db 1033 agggcaacacagctggtgtgatacagcaagcaagacgttcaaaaagcaagtgcaatc 1092  
QY 981 GGGACAAACAGATGGTGGCTTGTGATGTGAGAGCTTCAAAACCAAGGTCAAGCTATC 1040  
Db 1093 tgaagagatgagcagcgtgcgcgcgcgtgtatgagccctgcgcgcgcgcgcgcgcgcgcgc 1152  
QY 1041 TGAAGCAAGAGGAGACTGGGGGGGCAATGCTGTGGCACTGAGCTTGAATGACTTTGCCG 1100  
Db 1153 gctcctctctgc 1170  
QY 1101 GCTTCTCTCTGCAACAGG 1118

RESULT 2  
ID N81756 standard; DNA; 966 BP.

AC N81756;  
DT 19-OCT-1990 (first entry)  
DE Gene encoding polypeptide involved in protective mechanisms  
KW Immune response; cell growth; ss.  
FH Key  
FT CDS Location/Qualifiers  
FT /tag= a  
PN J63032898-A.  
PD 30-JAN-1988.  
PE 16-JUL-1986; 167518.  
PR 16-JUL-1986; JP-167518.  
PA (DAIN) Daiinippon Pharm KK.  
DR MPI; 88-068419/10.  
PT P-PSDB; P81342.  
PT New polypeptide and DNA encoding it - related to protective mechanisms  
PS such as immune response etc.  
PS Disclosure; P; Japanese.  
CC Polypeptide encoded is involved in protective mechanisms such as immune  
CC response, cell growth and activation of protective functions.  
SQ Sequence 966 BP; 238 A; 282 C; 247 G; 199 T;  
  
Query Match 8.1%; Score 133; DB 1; Length 966;  
Best Local Similarity 66.3%; Pred. No. 4,48e-71;  
Matches 270; Conservative 0; Mismatches 137; Indels 0; Gaps 0;  
  
Db 23 caggtctgtgtacccatcatcagcttgcacatataagcaaacatcacatgcaca 82  
QY 24 CAGGTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 83  
Db 83 acacacagctgtgtccagatccaggaagcgatgagatgagatgagatgagatgagatgagat 142  
QY 84 TCACCAACTGGGCCGATACAGACAGGGGGAGGCTCGCTTCTGCCCCAAGCACTTGAGCC 143  
Db 143 gcttcctgtgtacccatcatcagcttgcacatataagcaaacatcacatgcaca 202  
QY 144 CCAAGCTTTCACCCACCTATCTACGCCCTTCCGTGGCATGACCAACACACAGCTAGCA 203  
Db 203 cctggagatggaatgatagtacgctctacagcaltgctcaaacacatcaaacagcaacc 262  
QY 204 CCAAGCTTTCACCCACCTATCTACGCCCTTCCGTGGCATGACCAACACACAGCTAGCA 263  
Db 263 ccaacctgaaagactctctctgtctgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 322  
QY 264 CCAAGCTTTCACCCACCTATCTACGCCCTTCCGTGGCATGACCAACACACAGCTAGCA 323  
Db 323 agatagctctcaaacacacagagatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 382  
QY 324 ATATGTAAGCCAGGCCAACACCGTTCAGACTTTGCAACTGGCCATCAGGTTTCTGC 383  
Db 383 gcaacccatggtcttgatgagcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 429  
QY 384 GCAAAATACAGCTTTGACGGGCTTGACCTTGACTGGAGATGCCAGAA 430

RESULT 3  
ID Q90443 standard; cDNA to mRNA; 2504 BP.  
AC Q90443;  
DT 10-JAN-1996 (first entry)  
DE Murine oviduct specific glycoprotein cDNA.  
KW Murine oviduct specific glycoprotein; recombinant production;  
KW MOGP; ds.  
OS Mus musculus.  
FH Key  
FT CDS Location/Qualifiers  
FT /tag= a  
FT /note= "START codon absent"  
FT sig\_peptide 3..56  
FT /tag= b  
FT mat\_peptide 57..2156  
FT /tag= c  
PN J07107979-A.  
PD 25-APR-1995.  
PF 15-AUG-1994; 214227.



PR 19-AUG-1993: JP-227881.  
 PA (KINO-) KINOSER PEPTIDE KENKYUSHO KK.  
 DR WPI: 95-190179/25.  
 DR P-PSDB: R90443.  
 PT New DNA encoding an oviduct-specific glycoprotein - useful for  
 PT recombinant production in high quantities.  
 PS Claim 1: Pages 11-14; 22pp; Japanese.  
 CC 090443 encodes R73992 murine oviduct specific glycoprotein (MOGP).  
 CC The cDNA can be used for the commercial recombinant prodn. of  
 CC MOGP in high quantities.  
 SQ Sequence 2504 BP; 678 A; 616 C; 538 G; 672 T;

Query Match 5.7%; Score 93; DB 15; Length 2504;  
 Best Local Similarity 58.7%; Pred. No. 1.67e-43;  
 Matches 364; Conservative 0; Mismatches 253; Indels 3; Gaps 2;

```

Db 60 aaactgtgtctatctaccatctggcagcatcagtcgcaagccctgctccatcatg 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 68 AAACGTGCTGCTACTTACCACTGGCCCATACAGAGGGAGGCTCCCTCTG 127
Db 120 ccccatgacctggaacctctcttcttaagcatctgatatctgctcttgcataatgagc 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 CCCAAGAGACTTGACCCGACCTTTCACCCACCTCATCAGCTTCGCTGCATGACC 187
Db 180 aacatcagatgtgagcaagaatctgcagagatgaagaagctctctatccagatgcaac 239
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 AACACACAGCT-GAG--CACCATGAGTGAATGACGAGACTCTCTACAGGATTCAT 244
Db 240 aaactaaggaagaaagagagagagagagagagagagagagagagagagagagag 299
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 GGCCTGAGAGATGATGATCCCAAGCTGAAGACCTCTTACCATCGAGGCTGGAATTTC 304
Db 300 ggcacatcagctgacacgctatgctctccacactgtgcaaccgagaaagtattatgac 359
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 305 GGCACACAGAGATTCACAGATGATGATGACCCGACCAACACAGCTTGTCTGAC 364
Db 360 tcagatatactctctctgagatatactgctgagctgctgacatctctctctgac 419
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 365 TGGCCATAGATTTTGTGGCAATACAGTTTGAAGGCTTGAACCTTGAGCTGGAGTAC 424
Db 420 ccttgactagagagagagagagagagagagagagagagagagagagagagagagag 479
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 CCAGAGAGCAGAGGAGGAGCCCTGCGGTAGACAGAGAGAGAGAGAGAGAGAGAGAG 484
Db 480 ctccagctgctgctgagagagagagagagagagagagagagagagagagagagag 539
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 485 TTGGCCAAATGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
Db 540 gctgctgctctgagatcccaagcatcatcatcatcatcatcatcatcatcatcat 599
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 545 GCAAGCGGTTCAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
Db 600 aggcgtgtgattcatatcatatgttctctatgactatcatcatcatcatcatcatcat 659
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 605 CAGAACCTGATTTTGTCAACCTTATGAGGCTTCAAGACTTCCATGCTCTTGGAGAGAG 664
Db 660 acagagacacatagtcctct 679
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 ACGGACATACAGCCCTT 684

```

RESULT 4  
 ID Q10572 standard; DNA; 1047 BP.  
 AC Q10572;  
 DT 09-APR-1991 (first entry)  
 DE Human Natriuretic Peptide Receptor B.  
 KW NPB; ANP; CNP; kidney failure; heart failure; protein kinase;  
 KW hyperaldosteronism; glaucoma; guanylyl cyclase.  
 OS Homo sapiens.  
 FH Key  
 FH Peptide Location/Qualifiers  
 FT /label- signal sequence 1..22  
 FT Protein 12

```

FT /label- mature NPB
FT Domain 23..455
FT /label- extracellular domain
FT /note- "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label- transmembrane domain
FT Domain 479..1047
FT /label- cytoplasmic domain
FT /note- "GC and protein kinase activity"
FT Modified -site 24..26
FT /label- N-glycos-site
FT Modified -site 35..37
FT /label- N-glycos-site
FT Modified -site 161..163
FT /label- N-glycos-site
FT Modified -site 195..197
FT /label- N-glycos-site
FT Modified -site 244..246
FT /label- N-glycos-site
FT Modified -site 277..279
FT /label- N-glycos-site
FT Modified -site 349..351
FT /label- N-glycos-site
FT Modified -site 600..602
FT /label- N-glycos-site
FT W09100292-A.
PN 10-JAN-1991.
PD 22-JUN-1990; U03586.
PF 23-JUN-1989; US-370673.
PR (GERTH ) GENENTECH INC.
PA Chang M, Goeddel D, Lowe D;
PI WPI: 91-036711/05.
DR N-PSDB: Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3: Fig 1: 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPB analogues. The protein has a mol wt. of 115 kD (calculated Mr-
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

```

Query Match 5.6%; Score 92; DB 2; Length 1047;  
 Best Local Similarity 10.6%; Pred. No. 7.92e-43;  
 Matches 106; Conservative 279; Mismatches 603; Indels 15; Gaps 14;

```

Db 5 snnnuaanagvrmnagvntnnavnnnnhnsyavawrvnavanavanngrannvnd 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 GCCCAGTACAGACAGGGAGGAGCTGCTCCCTGCCCAAGAGATTGGACCCAGCTTGC 154
Db 65 rvssnnngacsnyannnsavdnknyhdnnngncvynnaas-varnasrvrnnntag 123
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 155 ACCCACCTCATCTACGCTTCCTGCGCATGACCAACACCACTGAGCAGCACTGAGTGG 214
Db 124 avagsnsakndyrtcnvrtgnsankngnvnvtnhnmwtaraannndardtdrnhnt 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 215 AATGACGAG-ACTCT-C-TACAGAGAGTTCAATGCGTGAAGAGATGATCCCAAGCTG 271
Db 184 nngvnanngsnsvnhvayrnngnnatnnnangrnyvncgnnnnnnnnnnan-r 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 272 AAGACCCCTTACCATCGAGGCTGGAGATTTCGGCACATCAAGTTCCACATATGTA 331
Db 243 nnnntgdyvnyndvngnsnrgnatratgrnvndrtlrnanannrnnantvnyrnn 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 GCCAGGCGCAACACCGTACAGACTTTCATCACTCGGCCCATCAGGTTCTCGGCAATAC 391
Db 303 nnnynnnnnnnnnrnnndvngnsmnnnngcnydgmnyanvnnntnnngtrndgn 362
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 392 AGCTTTGAGGGGCTTGACCTTGACTGAGGATACCCAGGAAGCCAGGAGGCC-TGCCGT 450
Db 363 rrvnkmgrryhyvtygnvmdkndrntdnywmgdngsdnnaahysganknwvtg 422
QY 451 AGACAAGAGAC-GCTTCAACAACCTGTACAGAGACTTGCCATCCCTTCCAGCAGAG 509
Db 423 rrrn-wv-kgsanndnncandndndscdklnstfnanvngtntmnygsnnnrk 480
QY 510 CCCAGACCTCAGGAGAGAGAGAGCCCTTCTCTAGTCAGAGCGGTTCACAGTGGCAGACT 569
Db 481 nmnnknaasmwrrnwnnnngnsnrhykagagrsntnsrsgssysnmtahykynnaant 540
QY 570 ATGTGATGTC-TGGATACAGAGGTGACAAATCCGCCAGAACCTTGATTTGTCAACTT 628
Db 541 ghnkgnvnanhkvokkrnnrttrnnnnkhmdvnnnbnhtrngscndnnncvtnycn 600
QY 629 -ATGGCCTTACGACTTCCATGGCTCTTGGGAGAGGTCAAGGACATAACAGCCCTCTTA 687
Db 601 rgsndnnndsnndnmrnyannndvkgmanhnsnshgnsksncvvsrnnkn 660
QY 688 CAAGAGGAGAGAGAGAGAGGTGTCAGAGAGCCGCTCAACGTGATGCTGTGTCAACA 747
Db 661 tdynasnrrstannndnnanyakknntannnnsgnnnttgmaadvyssngnnnnanr 720
QY 748 GTGGCTGAGAGAGGGGAGCCCTGCGCAGAGCTGATCCTTGGCATGCTACTACGAGAG 807
Db 721 sgnyngnngdnshknvknvkngrntrnyrnnsdrtnnnnnnvnmmrxcwandnarrndg 780
QY 808 CTCTCTTCACTACCTGCGCTCTCTATCAGACACCAAGTGGGGGCCCAAGCCACAGGCTTGG 867
Db 781 nkgmrnrxnkngrtsnndnnnnnnnnnyannnkhvnrtnaynnnkrkanannynnnh 840
QY 868 CACTCCAGGCCCTTACACCAAGAGAGAGGATGCTGGCTCTCTATGAAGTGTCTCTG 927
Db 841 svannnkngrtnvnanandsvtnyngsdvngntensanstmmvtrnndytcnd--and 898
QY 928 GAAGGGGCGCACCAAGAGATCCAGGATCCAGAGAGTGGCCCTCATCTTCCGGGACAA 987
Db 899 ndvykvtntgdaynvvsngnngnrrnhanmananndavssnnrrhrhndnnrrn 958
QY 988 CCAAGTGGGTGGCTTGTGATGATGTGGAGAGCTTCAAAAACCAAGGATGATCTGAAGCA 1047
Db 959 gvhtnvcagvnyvgnknrnyngndvntasrnsnngnanknhv 1001
QY 1048 GAAGGAGACTGGGCGG-GGCCATGATC-TGGGCATCTGAGACTTNG 1088

```

RESULT 5  
ID Q90442 standard; cDNA to mRNA; 1994 BP.  
AC Q90442;  
DT 10-JAN-1996 (first entry)  
DE Bovine oviduct specific glycoprotein cDNA.  
KW Bovine oviduct specific glycoprotein; recombinant production;  
KW BOP; ds.  
OS Bos taurus.  
FH Key  
FT CDS Location/Qualifiers  
FT /tag- a 2.1615  
FT /note- "START codon absent"  
FT sig\_peptide 2.55  
FT /tag- b  
FT mat\_peptide 56..1612  
FT /tag- c  
FT J07107979-A.  
PD 25-APR-1995.  
PF 15-AUG-1994; 214227.  
PR 19-AUG-1993; JP-227881.  
PA (KINO-) KINOSEI PEPTIDE KENKUYUSHO KK.  
DR MPI: 95-190179/25.  
P-PSDB: R73991.  
FT New DNA encoding an oviduct-specific glycoprotein - useful for  
FT recombinant protein production in high quantities.  
PS Claim 1; Pages 7-9; 22pp; Japanese.

```

CC Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC BOP in high quantities.
SQ Sequence 1994 BP; 445 A; 536 C; 520 G; 493 T;

Query Match
Best Local Similarity 58.1%; Pred. No. 4,11e-38;
Matches 360; Conservative 0; Mismatches 257; Indels 3; Gaps 2;

5.2%; Score 85; DB 15; Length 1994;

Db 59 aagctgtygttatttaccacactgggcatcagtcggccggccctgcctcaatcctg 118
QY 68 AAACGTGCTCTACTTACCAACTGGGCCCAATACAGACAGGAGGAGCTCGCTCTCTG 127
Db 119 cctgggagtcggagccggttctctcgcacccacactglatcttgcttgcctgtaggc 178
QY 128 CCAAGGACTTGGACCCAGGCTTGTGACCCACCTCATCTACGCTTGTGCTGGCATGACC 187
Db 179 aacaatcagattgttcctaaagatcccaagatgagaataatcctctaccagagttgaac 238
QY 188 AACCAACGCTGAGCACACACCTGAGTGGATACGAGAC--TC-TCTACAGGAAATTCAAT 244
Db 239 aagctcaagagagagacagagagggctgaaaacgctctgtccatcgcggggatggaactt 298
QY 245 GGCTGGAAGAGATGAATCCCAAGCTGAAGACCTGTAGGCATCGGAGCTGGAATTTC 304
Db 299 ggcacgctgaggttcacacacatgctgtccagctttccacacgggaaaggttcgtaagt 358
QY 305 GGCACCTGAAAGTTTCACGATATGTATGTAGCCAGGCCAACAACCTGTAGACCTTGTGTAAC 364
Db 359 tcaatgtagcgcctcctagagacacatgagctttagtgcctgagaccctctctcttctac 418
QY 365 TCGGCATCAGAGTTTCTGCCAAATATACGCTTTTACGCGCTGTACCTTGAGTGGAGATAC 424
Db 419 ccttgagctagaagcagccctgcgctgacccgctgagacccttgcttcttacttgaagag 478
QY 425 CCAGAAAGCCAGGGGAGGCGCTCGCTGAGACAAAGAGCGCTTCAACAACCTGTACAGAGAC 484
Db 479 ctccctgagcgccttcaagaatagagccagatcagcatgctccaaagctgctgtctct 538
QY 485 TTGGCAATGTGCTTCCACAGAGGAGCCAGCTCAGAGGAAGAAACGCTTCTTCTAGT 544
Db 539 gctgcgctctctgggagaccacatgctgcacgaagcgtatgaagcagcgcctctgggc 598
QY 545 GCAGGGGTTCACAGCTGGGCAACCTATGTGATCTGTGATCGAGTGGAGCAAAATGCC 604
Db 599 agaactcctgattcaatcagtgctgtctttagtacttaactgaagcctgggaaagctc 658
QY 605 CAGAACCTGGATTGTTGTCACACTTATGGCTTACGATCTCATGCTCTTGGGAGAGGTC 664
Db 659 aagagacacataagccctct 678
QY 665 ACGGACATTAACAGCCCTCT 684

```

RESULT 6  
ID Q10572 standard; DNA; 1047 BP.  
AC Q10572;  
DT 09-APR-1991 (first entry)  
DE Human Natriuretic Peptide Receptor B.  
KW NPB; ANP; BNP; CNP; Kidney failure; heart failure; protein kinase;  
KW hyperaldosteronism; glaucoma; guanyl cyclase.  
OS Homo sapiens.  
FH Key  
FT Peptide Location/Qualifiers  
FT /label- signal sequence 1..22  
FT Protein 12  
FT /label- mature NPB 23..455  
FT Domain  
FT /label- extracellular domain  
FT /note- "binds natriuretic peptides A, B and C"  
FT Domain 456..456  
FT /label- transmembrane domain 479..1047



Oy 68 AACTGTCCTGCTACTTCACCAACTGGGCCAGACAGACAGGGGAGGCGCTTCTCG 127  
Db 120 ccccgtagctgagatccctctctgtgtacacacatgatacttgccttgcctgagtgagc 179  
Oy 128 CCCAAGGACTTGGACCCCGCTTTCACCCCTCATCTTACGCCCTTCTGTCGACAGAC 187  
Db 180 aacaatcagatgtgttcacataatctccagatgagaataatctatccaaatcgaac 239  
Oy 188 AACCAACAGACTGAGCACCACCTGAG-TGA-ATGAGAGACTCTCTACACAGAGTTCAAT 244  
Db 240 aaacccaagagagagagacagagccctgaaacactactctgtgtgagagctggaactc 299  
Oy 245 GGCCTGAAGAAATGATGATCCCAAGCTGAGACCTGTTGACCATGCGAGCTGGAATTTTC 304  
Db 300 ggcacatccaggttcacacatgctgtccacccttgccagccgctgaaatatttggc 359  
Oy 305 GGCCTGAGAACTTTCAGATATGAGGACGACGCGCAACACCGTTCAGACTTTGTCAAC 364  
Db 360 teagttgtatctcctctggaacacatggtctgtgagctgtatctctctgttac 419  
Oy 365 TCGCCCATCAGGTTCTCGCAAAATACAGCTTGAAGCGCTTGACCTTGACTGAGAGTAC 424  
Db 420 cctgactacagagagagccc 439  
Oy 425 CCAGGAAGCCAGGGAGGCC 444

RESULT 8  
ID N81164 standard; DNA; 204 BP.  
AC N81164.  
DE 08-NOV-1990 (first entry)  
KW E.coli beta galactosidase alpha-fragment.  
OS Escherichia coli.  
FH Key location/Qualifiers  
FT misc-feature 19..69  
FT /\*tag- a  
FT /function-multiple cloning site  
FT primer\_bind 187..204  
FT /\*tag- b  
PN EP-285123-A.  
PD 05-MAY-1988.  
PF 30-MAR-1988; 105163.  
PR 03-APR-1987; US-034819.  
PA (SUSO) SUOMEN SOKERI OY.  
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;  
DR WPI: 88-279927/40.  
PT Introducing random point mutations into nucleic acids -  
PT by prepn of single stranded template, annealing a primer, elongation,  
PT misincorporation, completion of molecules and screening.  
PS Disclosure; P: English.  
CC Random point mutations were introduced into the alpha fragment of  
CC E.coli beta-galactosidase. The wild type sequence was obtained as a  
CC single stranded template and an oligonucleotide was hybridised to  
CC it to generate a popn of DNA molecules which terminate at all  
CC possible nucleotide positions within a specified region. The  
CC variable 3' ends generated in this way are used as primers for  
CC reverse transcriptase. Nucleotides are misincorporated by the  
CC transcriptase and the molecules are completed to forms that can be  
CC amplified and then expressed in a suitable host-vector system.  
CC The sequence covers all 176 diff base substitutions, most of which  
CC occurred singularly in any given mutant.  
CC See also P80575.  
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.1%; Score 51; DB 1; Length 204;  
Best Local Similarity 15.4%; Pred. No. 4.79e-16;  
Matches 18; Conservative 62; Mismatches 36; Indels 1; Gaps 1;  
Db 75 yycdchvgcgymrtthhrrmrbvdyrdynsdaaaycrrsvkycdcynachhd 134  
Oy 603 CCCAAGAACTGATTTGTCAACCTTATGCGCTACGACTTCATGCTCTTGAGAGAG 662

Db 135 hvvybbvynvnhnncocbnhvhcnvbnhnmhwayvhrdariddhcvchc 191  
Oy 663 TCACGGACATATACAGCCCTC-TACAAGAGCAAGAGAGAGTGTGACAGACC 718

RESULT 9  
ID Q51746 standard; cDNA; 91 BP.  
AC Q51746.  
DE 31-MAY-1994 (first entry)  
KW Oligonucleotide probe MK14-A  
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
OS Synthetic.  
PN EP-571911-A.  
PD 01-DEC-1993.  
PF 24-MAY-1993; 108325.  
PR 26-MAY-1992; US-889651.  
PA (BECT ) BECTON DICKINSON CO.  
PI Shank DD, Spears PA;  
DR WPI: 93-378844/48.  
PT New oligo:nucleotide probes specific for Mycobacteria - used for  
PT detection and amplification of Mycobacteria nucleic acid in  
PT samples  
PS Claim 3; Page 14; 23pp; English.  
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
CC cross reacted to a few non-mycobacterial spp. The probe may  
CC be useful as an initial screen for mycobacterial infection.  
CC See also Q51735-45 and Q51747-59.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.8%; Score 45; DB 9; Length 91;  
Best Local Similarity 2.0%; Pred. No. 2.21e-12;  
Matches 1; Conservative 47; Mismatches 3; Indels 0; Gaps 0;  
Db 10 gssvshyvvhvshhshvhhvshvvvvhvshvvhvshvhyvsv 60  
Cp 1361 GGCCTGCGGCGAGCTTGTGTAACAGCCGCCCGCTGCACAGCTGAGA 1311

RESULT 10  
ID Q51746 standard; cDNA; 91 BP.  
AC Q51746.  
DE 31-MAY-1994 (first entry)  
KW Oligonucleotide probe MK14-A  
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
OS Synthetic.  
PN EP-571911-A.  
PD 01-DEC-1993.  
PF 24-MAY-1993; 108325.  
PR 26-MAY-1992; US-889651.  
PA (BECT ) BECTON DICKINSON CO.  
PI Shank DD, Spears PA;  
DR WPI: 93-378844/48.  
PT New oligo:nucleotide probes specific for Mycobacteria - used for  
PT detection and amplification of Mycobacteria nucleic acid in  
PT samples  
PS Claim 3; Page 14; 23pp; English.  
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
CC cross reacted to a few non-mycobacterial spp. The probe may  
CC be useful as an initial screen for mycobacterial infection.  
CC See also Q51735-45 and Q51747-59.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.6%; Score 42; DB 9; Length 91;  
Best Local Similarity 3.8%; Pred. No. 1.37e-10;  
Matches 2; Conservative 45; Mismatches 5; Indels 0; Gaps 0;  
Db 12 svshyvvhvshhshvhhvshvvvvhvshvvhvshvhyvsvctc 63  
Oy 124 CCGCCCAAGACTTGAGCCCGCTTGCACCCAGCTTACCTCATCTACGCTTC 175

RESULT 11  
ID 070468 standard; DNA; 114 BP.  
AC 070468;  
DE Generic DNA sequence to generate a random TSAR peptide library.  
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
KW effector domain; concatenated heterofunctional protein; linker;  
KW direct; rapid; detection; screening; treatment; generic; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_feature 55..60  
FT /tag- a  
FT /note- "this sequence represents 'Z'; Z can be a  
FT sequence of 6, 9 or 12 nucleotides (see  
FT comments)"  
PN WO9418318-A.  
PD 18-AUG-1994.  
PF 01-FEB-1994; U00977.  
PR 01-FEB-1993; US-013416.  
PR 30-DEC-1993; US-176500.  
PR 31-JAN-1994; US-189331.  
PA (UNC-) UNIV NORTH CAROLINA.  
PI Fowlkes DM, Kay BK.  
DR WPI; 94-279739/34.  
DR P-PSDB; R65154.  
PT Identifying proteins or peptide(s) which bind a ligand - by  
PT screening a recombinant vector library expressing fusion proteins  
PT comprising a binding domain and an effector domain  
PS Disclosure; page 35; 255pp; English.  
CC 070468 is a generic DNA sequence used to generate random TSAR (Totally  
CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
CC represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X  
CC and Y are flanking restriction sites (X is not the same as Y) that are  
CC not specified further. Other generic sequences are shown in 070466-68.  
CC Other specific peptides generated by these generic sequences are shown in  
CC R5151-54. TSARs are concatenated heterofunctional proteins or peptides,  
CC comprising at least two functional regions - a binding domain with  
CC affinity for a ligand and a second effector peptide portion that is  
CC chemically or biologically active. They may further comprise a linker  
CC peptide between the 2 domains. The oligonucleotides are also designed so  
CC that the expressed peptide contains 2 or 4 cysteine residues positioned  
CC in, or flanking, the unpredicted or variant residues. These residues  
CC confer some degree of conformational rigidity to the peptides. The TSARs  
CC or compms. comprising a TSAR binding domain can be used in vivo to  
CC deliver a chemically or biologically active moiety, eg. metal ion,  
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the  
CC cell. They can also replace the function of macromolecules, eg.  
CC monoclonal or polyclonal antibodies and therefore circumvent the need  
CC for complex methods of hybridoma formation or in vivo antibody  
CC production. The TSARs are easily characterised and have designed activity  
CC allowing direct and rapid detection in a screening process.  
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;  
Query Match 2.3%; Score 38; DB 12; Length 114;  
Best Local Similarity 4.5%; Pred. No. 2.95e-08;  
Matches 5; Conservative 34; Mismatches 73; Indels 0; Gaps 0;

KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
KW effector domain; concatenated heterofunctional protein; linker;  
KW direct; rapid; detection; screening; treatment; generic; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_feature 55..60  
FT /tag- a  
FT /note- "this sequence represents 'Z'; Z can be a  
FT sequence of 6, 9 or 12 nucleotides (see  
FT comments)"  
PN WO9418318-A.  
PD 18-AUG-1994.  
PF 01-FEB-1994; U00977.  
PR 01-FEB-1993; US-013416.  
PR 30-DEC-1993; US-176500.  
PR 31-JAN-1994; US-189331.  
PA (UNC-) UNIV NORTH CAROLINA.  
PI Fowlkes DM, Kay BK.  
DR WPI; 94-279739/34.  
DR P-PSDB; R65153.  
PT Identifying proteins or peptide(s) which bind a ligand - by  
PT screening a recombinant vector library expressing fusion proteins  
PT comprising a binding domain and an effector domain  
PS Disclosure; page 35; 255pp; English.  
CC 070467 is a generic DNA sequence used to generate random TSAR (Totally  
CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
CC represented as follows: X(NNB)16(TGC)(NNB)12(TGC)(NNB)1Y. X  
CC and Y are flanking restriction sites (X is not the same as Y) that are  
CC not specified further. Other generic sequences are shown in 070466-68.  
CC Other specific peptides generated by these generic sequences are shown in  
CC R5151-54. TSARs are concatenated heterofunctional proteins or peptides,  
CC comprising at least two functional regions - a binding domain with  
CC affinity for a ligand and a second effector peptide portion that is  
CC chemically or biologically active. They may further comprise a linker  
CC peptide between the 2 domains. The oligonucleotides are also designed so  
CC that the expressed peptide contains 2 or 4 cysteine residues positioned  
CC in, or flanking, the unpredicted or variant residues. These residues  
CC confer some degree of conformational rigidity to the peptides. The TSARs  
CC or compms. comprising a TSAR binding domain can be used in vivo to  
CC deliver a chemically or biologically active moiety, eg. metal ion,  
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the  
CC cell. They can also replace the function of macromolecules, eg.  
CC monoclonal or polyclonal antibodies and therefore circumvent the need for  
CC complex methods of hybridoma formation or in vivo antibody production.  
CC The TSARs are easily characterised and have designed activity allowing  
CC direct and rapid detection in a screening process.  
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;  
Query Match 2.3%; Score 37; DB 12; Length 114;  
Best Local Similarity 4.8%; Pred. No. 1.11e-07;  
Matches 5; Conservative 32; Mismatches 68; Indels 0; Gaps 0;

Db 6 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnnnnnnnnn 65  
QY 109 GGGGAGGCTGCTTCTGCAAGAGCTTGAGCCCACTTGACACCACTCATCTA 168  
Db 66 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdb 110  
QY 169 GCGCTGCTGGCATGACCAACCAACGAGCTGAGACACACGAGTG 213

RESULT 13  
ID 070469 standard; DNA; 114 BP.  
AC 070469;  
DE Generic DNA sequence to generate a random TSAR peptide library.  
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
KW effector domain; concatenated heterofunctional protein; linker;  
KW direct; rapid; detection; screening; treatment; generic; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_feature 55..60  
FT /tag- a



in, or flanking, the unpredicted or variant residues. These residues confer some degree of conformational rigidity to the peptides. The TSARs or compens. comprising a TSAR binding domain can be used in vivo to deliver a chemically or biologically active moiety, eg. metal ion, radioisotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg. monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or in vivo antibody production. The TSARs are easily characterised and have designed activity allowing direct and rapid detection in a screening process.

Sequence 114 Bp; 0 A; 2 C; 2 G; 2 T;  
Query Match 2.2%; Score 36; DB 12; Length 114;  
Best Local Similarity 3.6%; Pred. No. 4.09e-07;  
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

Db 3 bnbdbnbdbnbdbnbtgcnbdbnbdbnbdbnbdbnbdbnbdbnnnnnn 62  
QY 130 CAAGGACTTGAGCCCGACGCTTGACCCACCTCATCTAGCCTCGCTGGCATGACCAA 189  
Db 63 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbtgcnbdbnbdb 114  
QY 190 CCACGAGCTGAGCACCACCTGAGTGAATGACGAGACTCTCTACCAGAGATTTC 241

Search completed: Sun Jun 29 19:58:47 1997  
Job time : 183 secs.







Db 183 -sgdyllqteydrhllgrtrdfnvsydlhgwsestghnspflslpdxkssa-fam 240  
Qy 185 PAGOYTY-DAAYEVDKLAQNIDFENLMAAYBFHGSWEKVTGHNHSPLEKXROEESGAASLNV 243  
Db 241 n--y--wrlgdpadskllmfgpavgrctfhllresknglgaasmgpaapgykvtqgflay 296  
Qy 244 DAAYQWQKQKTPHPSKLLIGMPYGRSTTLASSSDTRVGAAPNGSGTGPFTYEGMLAY 303  
Db 297 yevcsfllgraeekhwldhgyppayakgkewyddavsfsykamfvkkehfgamvwtllm 356  
Qy 304 YEVCSSW-KGAARKORIODKVPYIRFDNQWGFDEVESEFKTKVSLKQGLGGAMVALDL 362  
Db 357 ddvrgtfcgngppflvhlhneli 379  
Qy 363 DDFAGFSCNQGRYPLQTLRREL 385

	RESULT	2
ID	R73991	standard; Protein; 537 AA.
AC	R73991.	
DT	10-JAN-1996	(first entry)
DE	Bovine oviduct specific glycoprotein.	
KW	Bovine oviduct specific glycoprotein; recombinant production;	
KN	BOGP.	
OS	Bos taurus.	
FH	Key	Location/Qualifiers
FT	Peptide	1..18
FT	/label- sig-peptide	
FT	Peptide	19..537
FT	/label- mat-peptide	
PN	J07107979-A.	
PD	25-APR-1995.	
PF	15-AUG-1994; 214227.	
PR	19-AUG-1993; JP-227881.	
PA	(KINO-) KINOSEI PEPTIDE KENKYUSHO KK.	
DR	WP1. 95-190179/25.	
DR	N-PSDB; Q30442.	
PT	New DNA encoding an oviduct-specific glycoprotein - useful for	
PT	recombinant protein production in high quantities.	
PS	Claim 4; Pages 7-9; 22pp; Japanese.	
CC	O90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).	
CC	The cDNA can be used for the commercial recombinant prodn. of	
CC	BOGP in high quantities.	
SO	Sequence 537 AA;	

Query Match	39.1%	Score 1345	DB 14	Length 537
Best Local Similarity	50.4%	Pred. No. 6,48e-115		
Matches 191	Conservative	80	Mismatches 101	Indels 7
				Gaps 6
Db	6	q1ll1vk1hkhdsaaahk1vcyf1fwnafsrpdpas1lprldipfict1h1afafasmsnq4tvp	65	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	9	g6wllm1pwsaaklvocyf1fwnaqyrgogear1fepkdlpdlsc1h1ayafagm1h1olst	68	
Db	66	kdpqdek1lypefnk1kernrg1ktl1s1ggwnf1vtr1f1m1stf1sre1fvasvial1	125	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	69	1femndet-lyofefnclakmnp1k1tll1a1gcmnfetq1f1tmvat1annrc1fvas1refl	127	
Db	126	rhgffidg1d1ff1llyg1s1gspard1wt1v1f1leel1lg1fkneaq1t1m1p1r1l1saavsgd	185	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	128	k1ysfdg1d1d1m1e1y1g1s1g1s1p1a1v1d1e1r1t1t1v1o1d1a1n1f1o1q1e1a1o1t1s1g1k1e1r1l1l1s1a1a1v1p1a1g	187	
Db	186	ph1vckayear1l1gr1l1d1f1s1v1s1y1d1l1g1s1w1e1v1q1h1a1s1p1f1s-1-pqdpks-sa-yam	240	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	188	q1yv1dag1ev1d1k1a1o1n1d1f1v1n1m1a1d1f1g1s1w1k1y1t1g1h1s1p1l1k1r1o1e1s1g1a1a1s1l1w1da1v	247	
Db	241	nyw1rg1lpvpek1l1ng1b1t1y1tr1f1t1l1k1asq1e1l1a1g1v1a1g1s1p1k1y1t1k1e1q1f1a1y1e1c	300	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	248	qom1o1k1g1t1p1a1s1k1l1i1g1m1p1t1y1g1r1s1f1l1a1s1s1p1r1v1g1a1p1a1t1g1s1o1p1e1f1k1e1g1m1a1y1e1v1c	307	
Db	301	cfv1rrak1r1w1nd1y1v1p1a1f1k1g1k1ev1y1g1d1a1s1f1y1k1a1f1f1k1r1e1h1f1g1a1m1w1t1d1d1f1r	360	
	1	1	1	1
	1	1	1	1
	1	1	1	1
	1	1	1	1
Qy	308	sw-kat1o1r1o1d1o1k1v1p1i1f1r1o1m1o1w1g1d1d1v1e1s1f1t1k1y1s1k1o1k1g1a1m1w1a1l1d1d1f1a	366	

```

Db      361  gyfcgtgpfplvhtlnml 379
          | : | | : | : | | |
QY      367  GFSCNQGRYPPIQLRQEL 385

```

RESULT	3	
ID	R73992	standard; Protein; 718 AA.
AC	R73992;	
DT	10-JAN-1996	(first entry)
DE	Murine oviduct specific glycoprotein.	
KW	Murine oviduct specific glycoprotein; recombinant production;	
KW	MOGP.	
OS	Mus musculus.	
FH	key	Location/Qualifiers
FT	peptide	1..18
FT	/label- sig_peptide	
FT	peptide	19..718
FT	/label- mat_peptide	
PN	J07107979-A.	
PD	25-APR-1995.	
PF	15-AUG-1994; 214227.	
PR	19-AUG-1993; JP-227881.	
PA	(KINO-) KINOSER PEPTIDE KENKYUSHO KK.	
DR	WPI; 95-190179/25.	
DR	N-PSDB; Q90443.	
PT	New DNA encoding an oviduct-specific glycoprotein - useful for	
PT	recombinant protein production in high quantities.	
PS	Claim 4; Pages 11-14; 22pp; Japanese.	
CC	900443 encodes R73992 murine oviduct specific glycoprotein (MOGP).	
CC	The cDNA can be used for the commercial recombinant prodn. of	
CC	MOGP in high quantities.	
CC	Sequence 718 AA;	
CC		

	Query Match	38.7%	Score 1332;	DB 14,	Length 718;	
	Best Local Similarity	49.2%	Pred. No.	1,17e-113;		
	Matches 191;	Conservative	81;	Mismatches 107;	Indels 9;	Gaps 6;
Dd	7 lVllmkhsdaevklycyftmwnsrpgpsimphdiqfictlhlfafasnngiyak 66		:-		:-	
Oy	11 MVLKMPW-GSAKLVCYFTFMAQYRGGEARFLPKDLPSCHTLIYAFGMTNHOUSTT 69		:-		:-	
Dd	67 nlqdenvnyplsnfkLkerneleklIsIsgwngfgtsrtaflanrekfidsvlsflr 126	-				
Oy	70 EWNDET-LYGFENGKKNNPRLKTLIAIGNEFETQKTFTDMVAANNRQFFVSNAIFELR 128					
Dd	127 ihgfigdlffilypglyrgspshdtwrfifleeqlgferealltqbprillsaasgip 186					
Oy	129 KYSEFDGLDWEPGSGSPAYDKERFTTLVODLANAFQDAQSGKERLLLSAAPAQ 188					
Dd	167 slthsyaallgrldidinvlsydlnswsekftghnsplfsipedkska--y---am 241	: :::	:		:	
Oy	189 TYVDAGYEVDKIADNLDFVNLMAADPFHGSWEKVGHNSPLTKROEEGAASLVNDAAVQ 248					
Dd	242 ywrkgtpedklingfpdygrnflLkesknqIqtlaempaspkykkgagfiayeycs 301					
Oy	249 QWLKGRTSKSLILGMPTGRSFTLASSSDTRVAPARGSTGPFPFKEGMLATIEVCS 308					
Dd	302 fvygarkhmwidgyvypafkgekwLygdctdisfyxamykrrehfganwtlmdadvrg 361					
Oy	309 W-KGATRKRIDQKVPYPIFRDNQWVGDDVESFTKYSYLKQKGCGSAMWALDDDFAG 367					
Dd	362 tfcgnpfpilwhl-nellvgteasnsp 388					
Oy	368 FSCNGRGYRLQLTRLRELSTLYLRSGTF 395					
RESULT	4					
ID	P81342 standard; protein; 321 AA.					
AC	P81342:					
DT	19-OCT-1990 (first entry)					
DE	Polypeptide involved in protective mechanisms					
KM	Immune response; cell growth.					
PN	J63032898-A.					





QY 322 VPYIF-RD-NQWGFDDVESFKTKVSYLKQKGLGAMVW 358

RESULT 9  
ID R33072 standard; Protein: 423 AA.  
AC R33072:  
DE 30-JUN-1993 (first entry)  
KW Aphanocladium album pre-pro-endochitinase.  
OS Aphanocladium album.  
FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /label= signal  
FT /note= "pre-sequence"  
FT Peptide 23..34  
FT /note= "pro-sequence"  
FT Protein 35..423  
FT /label= endochitinase  
PN EP-531218-A.  
PD 10-MAR-1993.  
PF 04-SEP-1992; 402414.  
PR 06-SEP-1991; FR-011072.  
PA (SNFI ) ELF SANOFI.  
PA (ERAP ) SOC NAT ELF AOUTTAINE.  
PI Blaiseau P, Legoux R, Leguay J, Schneider M;  
DR WPI; 93-078700/10.  
DR N-PSDB; Q37711, Q37712.  
PT Recombinant DNA encoding fungal endo:chitinase - useful for  
PT protecting plants against fungi, insects, etc.  
PS Claim 1: Fig 1: 75pp; French.  
CC Clone CH3C was isolated from a cDNA bank prepared from A.album  
CC mRNA by screening with anti-chitinase antiserum. Three probes were  
CC designed based on the CH3C sequence, corresponding to the start,  
CC middle and downstream regions of the coding sequence. The probes  
CC were used to isolate the full-length coding sequence of A.album  
CC endochitinase from an A.album cDNA library in pTZ19R. A 1.6kb  
CC fragment was identified in several positive clones. The sequence  
CC of the fragment contained an ORF encoding a 423 amino acid protein.  
CC The mature endochitinase can be used to control pathogens  
CC such as fungi, bacteria, arthropods and nematodes on plants.  
SQ Sequence 423 AA;

Query Match 16.7%; Score 576; DB 6; Length 423;  
Best Local Similarity 36.2%; Pred. No. 2.20e-41;  
Matches 101; Conservative 65; Mismatches 95; Indels 18; Gaps 16;

Db 115 LKKQNRMKVMSIGGWTSW-nfpaaasaarctkfagsavgmfdgididweyp 173  
QY 83 LKGMNPKLTKLLAIGGMNFGTQKFTDMVATANNRQTFVNSAIRFLKRYSPDGLDMEYP 142  
Db 174 -adatatqamvlllgavrseldsyaay-akg-hhflislaaagpndynkklfa-e-1g 228  
QY 143 GSGSFAVVDKERTTLVQDLANFQDEAQTSGKERLLLSAAYAG-QTYVDAGYEVDKIA 201  
Db 229 kvldylnlaaydyagswnytdhdaanyaqpnaptr-yntdavgaylngvpankiv 287  
QY 202 QNIDFVNLAAYDPHSGMEKVTGHNSPLYKROESGAASLVNDAVQOMIQKGTSPASKLI 261  
Db 288 lmplygrsfq-qt--eg-1gkpynglgs-gsw--engldykalpk-agatvkcdttak 339  
QY 262 LGMPTGRSPTLLASSSDTRRGAPATSGTPPTKRGMLAYTEVCSWKGATKQRIDQK 321  
Db 340 gcysydpstkelisfdtpamistkswlykqglgsgmfw 378  
QY 322 VPYIF-RD-NQWGFDDVESFKTKVSYLKQKGLGAMVW 358

RESULT 10  
ID R33068 standard; Protein: 389 AA.  
AC R33068:  
DE 30-JUN-1993 (first entry)  
KW Aphanocladium album mature endochitinase.

KW filamentous fungus; pathogen resistance.  
OS Aphanocladium album.  
PN EP-531218-A.  
PD 10-MAR-1993.  
PF 04-SEP-1992; 402414.  
PR 06-SEP-1991; FR-011072.  
PA (SNFI ) ELF SANOFI.  
PA (ERAP ) SOC NAT ELF AOUTTAINE.  
PI Blaiseau P, Legoux R, Leguay J, Schneider M;  
DR WPI; 93-078700/10.  
DR N-PSDB; Q37704.  
PT Recombinant DNA encoding fungal endo:chitinase - useful for  
PT protecting plants against fungi, insects, etc.  
PS Claim 22: Page 60-61; 75pp; French.  
CC This mature protein from the filamentous fungus Aphanocladium  
CC album has endochitinase activity and is used to control pathogens  
CC such as fungi, bacteria, arthropods and nematodes on plants.  
CC Recombinant DNA comprising the endochitinase coding sequence is  
CC specifically intended for transforming Nicotiana tabacum, Helianthus  
CC annuus and Brassica napus to pathogen resistance.  
SQ Sequence 389 AA;

Query Match 16.5%; Score 569; DB 6; Length 389;  
Best Local Similarity 35.8%; Pred. No. 9.94e-41;  
Matches 100; Conservative 66; Mismatches 95; Indels 18; Gaps 16;

Db 81 LKKQNRMKVMSIGGWTSW-nfpaaasaarctkfagsavgmfdgididweyp 139  
QY 83 LKGMNPKLTKLLAIGGMNFGTQKFTDMVATANNRQTFVNSAIRFLKRYSPDGLDMEYP 142  
Db 140 -adatatqamvlllgavrseldsyaay-akg-hhflislaaagpndynkklfa-e-1g 194  
QY 143 GSGSFAVVDKERTTLVQDLANFQDEAQTSGKERLLLSAAYAG-QTYVDAGYEVDKIA 201  
Db 195 kvldylnlaaydyagswnytdhdaanyaqpnaptr-yntdavgaylngvpankiv 253  
QY 202 QNIDFVNLAAYDPHSGMEKVTGHNSPLYKROESGAASLVNDAVQOMIQKGTSPASKLI 261  
Db 254 lmplygrsfq-qt--eg-1gkpynglgs-gsw--engldykalpk-agatvkcdttak 305  
QY 262 LGMPTGRSPTLLASSSDTRRGAPATSGTPPTKRGMLAYTEVCSWKGATKQRIDQK 321  
Db 306 gcysydpstkelisfdtpamistkswlykqglgsgmfw 344  
QY 322 VPYIF-RD-NQWGFDDVESFKTKVSYLKQKGLGAMVW 358

RESULT 11  
ID W02159 standard; Protein: 866 AA.  
AC W02159:  
DE 14-JAN-1997 (first entry)  
KW Soluble chitinase.  
KW Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;  
KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;  
KW catabolism.  
OS Vibrio furnissii.  
PN W09625424-A1.  
PD 22-AUG-1996.  
PF 13-FEB-1996; U02332.  
PR 13-FEB-1995; US-386727.  
PA (UYDQ ) UNIV JOHNS HOPKINS.  
PI Bassler B, Chittlaru E, Keyhani N, Roseman S, Rowe C;  
PI Yu C;  
DR WPI; 96-393335/39.  
DR N-PSDB; T36390.  
PT Chitin biosynthetic enzymes end I, exo I and exo II - are  
PT periplasmic chito:dextrinase(s), periplasmic beta-glucanidase(s) and  
PT aryl beta-N-acetylglucosaminidase(s), respectively  
PS Example 4; Page 79-82; 101pp; English.  
CC Periplasmic chitodextrinase (W02156), periplasmic  
CC Beta-N-acetylglucosaminidase (W02157) and aryl  
CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
CC oligosaccharides with the structure (GlcNAc)<sub>n</sub> where n is 2 or



OY 227 PLYKROESGAASLVNDAVQWMLQKGPASKLILGMYTGRSTLASSSTRGAPAT 286  
 DB 463 hravxgtw--englydyqiaagfmsgewqlytdtaeapvfykfstgdltsddarsvq 520  
 OY 287 GSGTGPPTKEGMLAYEVCS-WKGAFTQ-RIQD-QKVPYIFR-DN-QWGEFDVESPK 341  
 DB 521 akgyvldkqjgglfsweldadn 543  
 OY 342 TKVSTLKQKGLGAMVWMLDLD 364

RESULT 14  
 ID R32547 standard; Protein: 238 AA.  
 AC R32547;  
 DT 17-JUN-1993 (first entry)  
 DE HAPV IE-1 gene protein (partial sequence).  
 KW Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;  
 KW Insecticide; immediate early-1.  
 OS Heliothis armigera nuclear polyhedrosis virus.  
 PN W09303144-A.  
 PD 18-FEB-1993.  
 PF 05-AUG-1992; AU0413.  
 PR 05-AUG-1991; AU-007576.  
 PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.  
 PI Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Srikantha A.  
 DR WPI: 93-076501/09.  
 DR N-PSDB: Q35982.  
 PT Insect virus with reduced capacity to occlude viral particles - used  
 PT for controlling proliferation of insect pests without horizontal  
 PT transmission.  
 PS Disclosure: Fig 2; 51pp; English.  
 CC The sequence is that encoded by the Heliothis armigera (Ha)  
 CC nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene  
 CC (partial sequence).  
 SQ Sequence 238 AA;

Query Match 7.9%; Score 271; DB 6; Length 238;  
 Best Local Similarity 34.1%; Pred. No. 1,65e-13;  
 Matches 58; Conservative 43; Mismatches 51; Indels 18; Gaps 14;  
 DB 44 erdkflsppgyreflytwkf-ftdgididwefpgkganpnvgdverndntyalldgela 102  
 OY 115 NMQITVNSAIR-R-F-L-KRYSFDGLDWEYPSQGS-PAV-DKER---FTTLVODLAN 164  
 DB 103 mldq-vqigturtltaitsagidkiaav-ncdraqyldkifmysydfkxgawntdig 160  
 OY 165 AFQQAQTSGRKRLLSAAVPAQGYVDAGYEVDKIAQNLDFVNLMAVDFHGSWEKVT-G 223  
 DB 161 hgtaly--gsawknepytaavavdallaqvnp-kkivlgvamygrgwt 207  
 OY 224 HNSPLYKROESGAASLVNDAVQWML-QKCTPASKLILGMYTGRSPT 272

RESULT 15  
 ID W02156 standard; Protein: 1046 AA.  
 AC W02156;  
 DT 14-JAN-1997 (first entry)  
 DE Periplasmic chitodextrinase.  
 KW Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;  
 KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;  
 KW catabolism.  
 OS Vldrio furnisii.  
 PN W09625424-A1.  
 PD 22-AUG-1996.  
 PF 13-FEB-1996; U02332.  
 PR 13-FEB-1995; US-386727.  
 PA (UJDO ) UNITV JOHNS HOPKINS.  
 PI Baessler B, Chittaru E, Keyhani N, Roseman S, Rowe C;  
 PI Yu C;  
 DR WPI: 96-393335/39.  
 DR N-PSDB: T36387.  
 PT Chitin biosynthetic enzymes end I, exo I and exo II - are

PT periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and  
 PT aryl beta-N-acetylgluco:amidase(s), respectively  
 PS Claim 2; Page 68-71; 101pp; English.  
 CC Periplasmic chitodextrinase (W02156), periplasmic  
 CC Beta-N-acetylglucosaminidase (W02157) and aryl  
 CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
 CC oligosaccharides with the structure (GlcNAc)<sub>n</sub> where n is 2 or  
 CC higher, by contacting them with soluble chitin. The enzymes are  
 CC encoded by the genes endI, exoI and exoII respectively. They are  
 CC all genes involved in the catabolic pathway of chitin.  
 SQ Sequence 1046 AA;

Query Match 6.4%; Score 219; DB 19; Length 1046;  
 Best Local Similarity 33.3%; Pred. No. 5.16e-09;  
 Matches 48; Conservative 40; Mismatches 42; Indels 14; Gaps 10;  
 DB 454 etfadsavemrkxrfdgldidrlsridggtqnpddafesarraylmsyhelmylr 513  
 OY 117 QTFVNSAIRFLRKIFFDLDD-W--EYPSQGSRA-VD-KE-RTTLY---QDLANFQ 167  
 DB 514 exldvasagdvhymltaaps-ayllrmetnavtgyldvnylmsydlngavndhvg 572  
 OY 168 QBAQ-TSGKE--RLLSAAVPAQGYVDAGYEVDKIAQNLDFVNLMAVDFHGSWEKVTGH 224  
 DB 573 naalydtgkdselaqv-nvygtaq 595  
 OY 225 NSPLYKROESGAASLVNDAVQ 248

Search completed: Fri Jun 27 16:54:04 1997  
 Job time : 116 secs.





\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
MARCH 21 1997  
\*\*\*\*\*

\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*

March\_dp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 27 16:49:12 1997; Maspar time 21.48 seconds  
Tabular output not generated. 618.601 Million cell updates/sec

Title: >US-08-663-618A-2  
Description: (1-466) from US08663618A.pep  
Perfect Score: 3443  
Sequence: 1 MVRSVAMGFWLMLPMGWS.....QQSCPTGLVFNCSCKCCTWN 466

Scoring table:  
PAM 150  
Gap 11

Searched: 89912 segs, 28507787 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r50  
1:ann1 2:ann2 3:ann3 4:ann4 5:annan1 6:annan2 7:annan3  
8:annan4 9:annan5 10:annan6 11:annan7 12:annan8  
13:annan9 14:annan10 15:unrec 16:unrev

Statistics: Mean 48.408; Variance 102.872; scale 0.471

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1519	44.1	383	13	A49562 cartilage glycoprotein	8.84e-265
2	1462	42.5	383	16	heparin-binding glyco	1.71e-253
3	1459	42.4	381	14	BRP39 protein - mous	6.73e-253
4	1459	42.4	381	14	BRP39 protein - mous	6.73e-253
5	1380	40.1	399	5	secretory protein YM	2.80e-237
6	1361	39.5	539	14	estrogen dependent o	1.58e-233
7	1345	39.1	537	14	oviduct-specific gly	2.27e-230
8	1185	34.4	483	12	chitinase (EC 3.2.1.	1.85e-179
9	1086	31.5	504	12	chitinase (EC 3.2.1.	1.85e-179
10	1031	29.9	554	12	chitinase (EC 3.2.1.	1.04e-168
11	698	20.3	699	10	chitinase (EC 3.2.1.	2.46e-104
12	594	17.3	378	5	chitinase (EC 3.2.1.	1.33e-84
13	582	16.9	424	5	chitinase (EC 3.2.1.	2.42e-82
14	576	16.7	423	5	chitinase (EC 3.2.1.	3.26e-81
15	561	16.3	427	12	chitinase (EC 3.2.1.	2.14e-78
16	546	15.9	427	12	chitinase (EC 3.2.1.	1.38e-75
17	446	13.0	563	16	chitinase precursor	4.70e-57
18	404	11.7	561	9	chitinase (EC 3.2.1.	2.11e-49
19	344	10.0	499	9	chitinase (EC 3.2.1.	1.19e-38
20	345	10.0	799	8	chitinase (EC 3.2.1.	7.93e-39
21	344	10.0	820	16	chitinase (EC 3.2.1.	1.19e-38

22	337	9.8	499	9	S04856	chitinase (EC 3.2.1.	2.06e-37
23	314	9.1	511	11	S61156	probable membrane pr	2.24e-33
24	287	8.3	610	10	JH0573	chitinase (EC 3.2.1. <th>1.06e-28</th>	1.06e-28
25	266	7.7	452	12	JC4038	47K glycoprotein - f <th>4.02e-25</th>	4.02e-25
26	263	7.6	1146	11	S07915	R22 protein - yeast <th>1.29e-24</th>	1.29e-24
27	246	7.1	413	5	JC2135	chitinase (EC 3.2.1. <th>9.11e-22</th>	9.11e-22
28	229	6.7	124	5	S57715	chitinase (EC 3.2.1. <th>5.81e-19</th>	5.81e-19
29	172	5.0	36	14	A27682	chitinase (EC 3.2.1. <th>5.52e-10</th>	5.52e-10
30	172	5.0	83	13	A57954	33K whey protein - b <th>5.52e-10</th>	5.52e-10
31	136	4.0	29	14	S57204	estradiol-stimulated <th>5.52e-10</th>	5.52e-10
32	136	4.0	29	14	S55693	oviduct-specific gly <th>8.32e-05</th>	8.32e-05
33	121	3.5	597	5	S33848	oviduct-specific sta <th>8.32e-05</th>	8.32e-05
34	121	3.5	597	5	S32039	chitinase (EC 3.2.1. <th>8.07e-03</th>	8.07e-03
35	112	3.3	642	16	S37623	chitinase (EC 3.2.1. <th>8.07e-03</th>	8.07e-03
36	110	3.2	408	10	A41961	homologues with orf <th>1.08e-01</th>	1.08e-01
37	104	3.0	385	9	S28440	chitinase (EC 3.2.1. <th>1.90e-01</th>	1.90e-01
38	104	3.0	422	4	ZWEC	probable coproporphy <th>9.80e-01</th>	9.80e-01
39	102	3.0	662	11	S61193	unuc protein - Esche <th>9.80e-01</th>	9.80e-01
40	105	3.0	787	8	PC1232	probable membrane pr <th>1.67e+00</th>	1.67e+00
41	105	3.0	1409	4	OFECF	retrovirus-related c <th>7.49e-01</th>	7.49e-01
42	101	2.9	562	3	KRHDB	retrovirus-related p <th>7.49e-01</th>	7.49e-01
43	101	2.9	958	1	JN0102	keratin, 56K type II <th>2.17e+00</th>	2.17e+00
44	101	2.9	1045	2	SUSEM	glucan 1,4-alpha-glu <th>2.17e+00</th>	2.17e+00
45	101	2.9	1070	11	S19686	serine proteinase (E <th>2.17e+00</th>	2.17e+00
						alpha-glucosidase (E <th>2.17e+00</th>	2.17e+00

## ALIGNMENTS

RESULT	ENTRY	1	ALIGNMENTS
	A49562	#type complete	
	cartilage glycoprotein gp39 precursor - human		
	39K synovial protein		
	ALTERNATE NAMES		
	ORGANISM		
	DATE		
	23-Mar-1995	#sequence	revision 23-Mar-1995
	03-May-1996	#text	change

ACCESSIONS	REFERENCE
A49562; S10677; A33162	
A49562	Hakala, B.E.; White, C.; Recklies, A.D.
	J. Biol. Chem. (1993) 268:25803-25810
	Human cartilage gp-39, a major secretory product of articular
	chondrocytes and synovial cells, is a mammalian member of a
	chitinase protein family.

#accession	A49562
#status	preliminary
#molecule_type	mRNA
#residues	1-383 #label HAK
#cross-references	GB:M80927
REFERENCE	S10677
#authors	Nyirkos, P.; Golds, E.E.
#journal	Biochem. J. (1990) 269:265-268
#title	Human synovial cells secrete a 39 kDa protein similar to a
	bovine mammary protein expressed during the non-lactating
	period.

#cross-references	MUTD:90328983
#accession	S10677
#molecule_type	protein
#residues	22-40, 'X', 42-45 #label NY2
	cartilage; extracellular protein; glycoprotein

KEYWORDS	FEATURE
1-21	#domain signal sequence #status predicted #label SIG
22-383	#product cartilage glycoprotein gp39 #status predicted
	#label MAT

SUMMARY	#length	383	#molecular-weight	42613	#checksum	2942
---------	---------	-----	-------------------	-------	-----------	------

Query Match	44.1%	Score 1519	DB 13	Length 383
Best Local Similarity	52.8%	Pred. No. 8.84e-265		
Matches	201	Conservative	88	Mismatches 83
		Indels	9	Gaps 5

Db	8	tgfvvlllgccsasklycytswsgregdscfpalgrfcthtlysfansndhd 67
Oy <td>8<td>AGFVFLMIPWMSAKKVCFTMAQYRGEARLPKDDPSCTHTLYFAFAGTNNQLS 67</td></td>	8 <td>AGFVFLMIPWMSAKKVCFTMAQYRGEARLPKDDPSCTHTLYFAFAGTNNQLS 67</td>	AGFVFLMIPWMSAKKVCFTMAQYRGEARLPKDDPSCTHTLYFAFAGTNNQLS 67
Db <td>68<td>twewndvlygmintknrpnltllsvgwmfgsgqrfisklansltqsrfftkvppfl 127</td></td>	68 <td>twewndvlygmintknrpnltllsvgwmfgsgqrfisklansltqsrfftkvppfl 127</td>	twewndvlygmintknrpnltllsvgwmfgsgqrfisklansltqsrfftkvppfl 127

QY	68	TTEWDELTYEFGGLKMNKLTLLAIGWNGCTKFTDWMATANNRPTVNSAIRFL	127
Db	128	rtfhfgdglawlypgrr-----dkqhftllkemkaefikeag-pgkqgllsaalsag	181
QY	128	RKYSFSDGIDLMEWEPGSGSPAVDKERFTTLVODLANAFQOEAQTSKERILLISAANPAG	187
Db	182	kvtldssydaiksgisqldffislmcyddfhgawrgtctghspflsfqgdeasdrtsencdyav	241
QY	188	QTYVDAGGEVDKIAQNIDFVNLMAVDHSGSEKVTGNSPLRYKROESGGAASLNDAAV	247
Db	242	gymrlrgapacklmgjptfgrftclass-etvgapibsgpgjprgtkeagctlayeic	300
QY	248	QOWLOKGRPAKLLIGMPTYGRSTLTASSSDTRKGAATSGTPGPTKGGHALAYEVC	307
Db	301	dfligatvhtlrgqvyatkqnvwyddqesvkskvyjldkrlqaglamvwaljldfdg	360
QY	308	SW-KGATKQRIQDQKVPYIFRDNQWGVDPVESKTKVSYLKKQKGLGAGAVMALDDDF	366
Db	361	gsfgqgdrlrflptalkdala	361
QY	367	GFSCWOG-RYPLIQTLRQELS	366
RESULT	2		
ENTRY	5S1327	#type complete	
TITLE	heparin-binding glycoprotein 38k - pig		
ORGANISM	#format_name Sus scrofa domestica #common_name domestic pig		
DATE	07-May-1995 #sequence_revision 07-May-1995 #text_change 07-May-1995		
ACCESSIONS	5S1327		
REFERENCE	5S1327		
Authors	Shackelton, L.M.; Mann, D.M.; Mills, A.J.T.		
#submitter	Submitted to the EMBL Data Library, January 1995		
#description	Identification of a 38kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling.		
#accession	5S1327		
#status	preliminary		
##residues	1-383 #label SHA		
##cros-references	EMBL:Z47603		
SUMMARY	#length 383 #molecular-weight 42443 #checksum 2907		
Query Match	42.5%; Score 1462; DB 16; Length 383;		
Best Local Similarity	51.0%; Pred. No. 1.71e-253;		
Matches	189; Conservative 87; Mismatches 94; Indels 9; Gaps 5;		
Db	1	mglyraetqgfayalvlgscaaayklycvytsvsgyredgdscfpdaipfclthlysfan	60
QY	1	WVRSYAMAGFVNLIMPGSAAKLVCYFTTNAOYRQGEARLPKDDPSICTHLIYAFAG	60
Db	61	isnaeidclenmdvlydntlntklnrnpnlklltsvsgwmfsgsfkiasntgsrftfl	120
QY	61	MTNHLSTTEWDELTYEFGGLKMNKLTLLAIGWNGCTKFTDWMATANNRPTV	120
Db	121	ksvpfllrthfgdglawlypgrr-----dkrhlrtlvkemkaefyreal-pyterlll	174
QY	121	NSAIRFLRKYSFDGIDLMEWEPGSGSPAVDKERFTTLVODLANAFQOEAQTSKERILL	180
Db	175	sgavsagkvaaidrydiaqisqnlidflslydfdhgawrgtctghspflsfqgdeasdrf	234
QY	181	SAAPVAGQTYVDAGGEVDKIAQNIDFVNLMAVDHSGSEKVTGNSPLRYKROESGGAAS	240
Db	235	snadgavsvyrlrgapanlkmjptfgrftclassk-tvqgaaaspgjprgtkeqyl	293
QY	241	LNVDAVOOVLQKGRPAKLLIGMPTYGRSTLTASSSDTRKGAATSGTPGPTKGGM	300
Db	294	layeyicdflgagatvrlpgqvyatkqnvwyddqesvkskakyjksrqqlaglamvwt	353
QY	301	LAYEVCW-SW-KGATKQRIQDQKVPYIFRDNQWGVDPVESKTKVSYLKKQKGLGAGAVMA	359
Db	354	ldlddfrnfcognrlrflptalsakdvla	361

QY	360	LDLDDPAGFSCNQG-RYPLIOTLRQELS	386						
RESULT	3								
ENTRY		S61550	#type complete						
TITLE		BRP39 protein - mouse							
ORGANISM		#formal_name Mus musculus	#common_name house mouse						
DATE		10-Apr-1996	#sequence_revision 19-Apr-1996	#text_change					
		03-May-1996							
ACCESSIONS		S61550							
REFERENCE		S61550							
#authors		Morrison, B.W.							
#submission		submitted to the EMBL Data Library, November 1995							
#accession		S61550							
#status		preliminary							
#molecule_type		mRNA							
#residues		1-381	#label MOR						
#cross-references		EMBL:X93035							
SUMMARY		#length 381	#molecular_weight 43001	#checksum 5107					
Query Match		42.4%	Score 1459;	DB 14;	Length 381;				
Best Local Similarity		49.9%	Pred. No. 6,73e-253;						
Matches 194;	Conservative	88;	Mismatches 96;	Indels 11;	Gaps 9;				
Db	1	mgmraalgfavlmlqscsakyklycfyswsgregvgsflpdaipfclthlysfan	60						
QY	1	MVRSVAMAGFVILMLIPMGCSAKLVCYFNMAQVROGEARFLPKDDPSLCTHLIYAFAG	60						
Db	61	issdmstswendensydklnkiktrtnlktllsvsgwkfgexrfseasntertaf	120						
QY	61	MT-NHQLSTTEWDEFLVQEFNGLKMKMKLTLTLAIGMNFQGTOKFTMVATVANNRQTE	119						
Db	121	vsrvapflrsyfgdglawlyp-r-lr-dkyfstlikelnaefkveq-pgrekll	174						
QY	120	VMSAIRFLRKYSFDDGLDMEXPGSGSAVNDKERITTVIQLDIANFQGEAQSGBRLL	179						
Db	175	lsaaalsagkvaicdyddlaqiaqlidflnlttydfhngvrqitlghspjlfqgkdxtrfd	234						
QY	180	LSAAVPAQGTVDAGYEVDKRIQONLDFVIMAYDFHFGSWEKVTGHSNPLYKKRQESGAA	239						
Db	235	yshvnyavaymrlrqaqaekllmjpdtfgskftlaas-engdgapisgeglpgrftkeag	293						
QY	240	SLNVDAVQWQLOKGPAPKLLIGMTYRSFTLASSDTRGAPATVGGTPEPFTKEGG	299						
Db	294	tlayeicdflgaaethrisnekvpatkqngvvygvehkesvknkygflkekklagawv	353						
QY	300	MLAYVEVCWM-KGATQKRIQDDQKVPRIIFRDNDONGVDVDESEFTKYSYLKQKGLGAMW	358						
Db	354	aiddldfgg-tcqpkeffiplnaikaala	381						
QY	359	ALDLDDPAGFSCNQGRY-PLIOTLRQELS	386						
RESULT	4								
ENTRY		I48271	#type complete						
TITLE		BRP39 protein - mouse							
ORGANISM		#formal_name Mus musculus	#common_name house mouse						
DATE		02-Jul-1996	#sequence_revision 02-Jul-1996	#text_change					
		02-Jul-1996							
ACCESSIONS		I48271							
REFERENCE		I48271							
#authors		Morrison, B.W.; Leder, P.							
#journal		Oncogene (1994) 9:3417-3426							
#title		neu and ras initiate murine mammary tumors that share genetic markers generally absent in c-myc and int-2-initiated tumors.							
#cross-references		MUID:95060797							
#accession		I48271							
#status		preliminary;	translated from GB/EMBL/DBJ						
#molecule_type		mRNA							
#residues		1-381	#label RES						
#cross-references		EMBL:X93035;	NID:g1085065;	CDS_PTD:g1085066					

GENETICS  
#note  
SUMMARY

gene name brp39  
#length 381 #molecular-weight 43001 #checksum 5107

Query Match 42.4%; Score 1459; DB 14; Length 381;  
Best Local Similarity 49.9%; Pred. No. 6,73e-253;  
Matches 194; Conservative 88; Mismatches 96; Indels 11; Gaps 9;

Db 1 mgmraatgafavlmllqscgahlycyfswgyregvgsflpdaipfictllysfan 60  
1 mrsavmagfvmllmipwgsaaklvcyftmwaoyrgoeaflpkdpslcthllyafag 60  
Qy 1 mrsavmagfvmllmipwgsaaklvcyftmwaoyrgoeaflpkdpslcthllyafag 60  
Db 61 isadmlstcwmdesnydklnktrtnlklisvgyqkfgkrfseasnterraf 120  
Qy 61 mt-nholsttewndetlyoeefnlgkkmprkltilaigwnfctokftdmvatanrort 119  
Db 121 vsavapflrsygfddglawlyp-r-lr--dkyfsclikeknaefkevq-pgarekl 174  
Qy 120 vnsaifrlkrkysfddglmdexyrgsgspavdkerrfttlvoodlanafogeaotsgerell 179  
Db 175 lsaalsagkvaldtygdiaqiaqhldfinlmtydfhgwvrtqthspilfgqkdrtrfd 234  
Qy 180 lsaavpagotyydagyeavkiaonldfvmilmaydfhgswekvtghnsplykrqeesgaaa 239  
Db 235 ysnvnyavgymlrlgagaskllmgipthgkfstlass-englqapslsgelipgftkeag 293  
Qy 240 slnvdaavoomokgrtpaskllgmptyrgrsftlasssdrvgapartsgtppgpfkeeg 299  
Db 294 clayaelcdfllygaevhrylnekvfakgnygvgyehkvsnkvgflkekklagawmv 353  
Qy 300 mlayevvcswm-kgatkoriodokvpyifrdnqmwgfdvdesfktvksylkqkglagawmv 358  
Db 354 aldidddfg-tcqpkefffltnaikdala 381  
Qy 359 alldlddfagfscnqgry-pliotrlrqels 386

RESULT 5  
ENTRY S27879 #type complete  
TITLE secretory protein YM-1 precursor - mouse  
ORGANISM #normal\_name musculus #common\_name house mouse  
DATE 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 03-May-1996

ACCESSIONS S27879  
REFERENCE S27879  
#authors Chang, N.C.A.; Liu, C.H.; Chang, A.C.  
#submission submitted to the EMBL Data Library, June 1992  
#description Molecular characterization of a secretory protein (YM-1) transiently expressed by activated murine peritoneal macrophages.

#accession S27879  
#molecule\_type mRNA  
#residues 1-399 #label CHA  
#cross-references EMBL:M94584  
CLASSIFICATION #superfamily streptomyces chitinase chl40  
FEATURE  
1-21 #domain signal sequence #status predicted #label SIG  
22-399 #product secretory protein YM-1 #status predicted #label MAT

SUMMARY #length 399 #molecular-weight 44528 #checksum 3208

Query Match 40.1%; Score 1380; DB 5; Length 399;  
Best Local Similarity 46.4%; Pred. No. 2.80e-237;  
Matches 181; Conservative 94; Mismatches 112; Indels 3; Gaps 3;

Db 121 lgvairrlrgynfdglnldqygsrsgspkdkhlfsylvkemrkafeesvekdiprll 180  
Qy 120 vnsaifrlkrkysfddglmdexyrgsgspavdkerrfttlvoodlanafogeaotsgerell 179  
Db 181 ltsagaidlvksgrtsrncslsldyigmtlydldhdpkdytgensplykspdydiksa 240  
Qy 180 lsaavpagotyydagyeavkiaonldfvmilmaydfhgswekvtghnsplykrqeesgaaa 239  
Db 241 dlnvdsllsywkdhaaseklvfgfpaqyhtllsdpstktyigapltstgppgkydeeg 300  
Qy 240 slnvdaavoomokgrtpaskllgmptyrgrsftlasssdrvgapartsgtppgpfkeeg 299  
Db 301 llayevvcfllegatvewdpagvpyaygngewgyddvrsfklkagylkdmnlgavv 360  
Qy 300 mlayevvcswm-k-gatkoriodokvpyifrdnqmwgfdvdesfktvksylkqkglagawmv 357  
Db 361 wplmdddfsgrfchgrhfrpitsclkgdlnl 390  
Qy 358 waldddfagfscnqgrypliotrlrqelsl 387

RESULT 6  
ENTRY I46470 #type complete  
TITLE estrogen dependent oviduct protein precursor - sheep  
ORGANISM #normal\_name ovis orientalis aries, Ovis ammon aries  
DATE 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Aug-1996

ACCESSIONS I46470  
REFERENCE I46470  
#authors Desouza, M.M.; Murray, M.K.  
#journal Endocrinology (1995) 136:2485-2496  
#title An estrogen-dependent secretory protein, which shares identity with chitinases, is expressed in a temporally and regionally specific manner in the sheep oviduct at the time of fertilization and embryo development.

#cross-references MUID:95269691  
#accession I46470  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type mRNA  
#residues 1-539 #label DES  
#cross-references EMBL:u16719; NID:q9885600; CDS:PID:q9885601  
SUMMARY #length 539 #molecular-weight 59535 #checksum 2120

Query Match 39.5%; Score 1361; DB 14; Length 539;  
Best Local Similarity 48.8%; Pred. No. 1.58e-233;  
Matches 189; Conservative 85; Mismatches 106; Indels 7; Gaps 5;

Db 1 mgklllvglmlmklhgdgaahklycyfenvafsrpgasllprldpflchlyvafas 60  
Qy 1 mrsavmagfvmllmipwgsaaklvcyftmwaoyrgoeaflpkdpslcthllyafag 60  
Db 61 mnnngivpdkpideklltyefnklkernnglktllisvsgwnfgtsrftmstfsnref 120  
Qy 61 mnnholsttewndetlyoeefnlgkkmprkltilaigwnfctokftdmvatanrort 119  
Db 121 vksviallrlthgfdgldlfllypqlrsgpdrwtvflleelqafkneagltmrrll 180  
Qy 120 vnsaifrlkrkysfddglmdexyrgsgspavdkerrfttlvoodlanafogeaotsgerell 179  
Db 181 lsaavsgdphvlgkaydalllgrlldfslvlsydlhgswekvtghnspilfsipgdkksa 240  
Qy 180 lsaavpagotyydagyeavkiaonldfvmilmaydfhgswekvtghnsplykrqeesgaaa 239  
Db 241 -yam--sy--wrglgyvpeklimgiplygrthllrasqnelagagaagapgytkyqg 295  
Qy 240 slnvdaavoomokgrtpaskllgmptyrgrsftlasssdrvgapartsgtppgpfkeeg 299  
Db 296 flayevvcsfvgtrakrwnldgvyvpyafkgykewvyddalsfygkafikrehfggawv 355  
Qy 300 mlayevvcswm-kgatkoriodokvpyifrdnqmwgfdvdesfktvksylkqkglagawmv 358  
Db 356 tldiddfrngfcgtpfpahntlmnl 382

OY	359	AADDLDFAGFSCNGKRYPLIQTIRQEL	385
RESULT	7	S57197	#type fragment
ENTRY		oviduct-specific glycoprotein	95k precursor - bovine
TITLE		(fragment)	
ORGANISM		#normal_name Bos primigenius taurus	#common_name cattle
DATE		28-Oct-1995	#sequence_revision 03-Nov-1995 #text_change 03-May-1996
ACCESSIONS		S57197	
REFERENCE		S57197	
#authors		Sendai, Y.; Abe, H.; Kikuchi, M.; Satoh, T.; Hoshi, H.	
#journal		Biol. Reprod. (1994) 50:927-934	
#title		Purification and molecular cloning of bovine oviduct-specific glycoprotein.	
#accession		S57197	
##status		preliminary	
##molecule_type		mRNA	
##residues		1-537	##label SEN
##cross-references		EMBL:D16639	
SUMMARY		#length 537	#checksum 1695
Query Match		39.1%	Score 1345; DB 14; Length 537;
Best Local Similarity		50.4%	Pred. No. 2.27e-230;
Matches	191;	Conservative	80; Mismatches 101; Indels 7; Gaps 6;
Db	6	glllvlkhhgaahklvcyftnwaafsrppasllprldbpflcthyfafamsnglvp	65
OY	9	GFWLLIMPMGSAKLVCYFTNWAQYRQGEARFLPRDLDPSTCHILTYAFAGMTNHQLST	68
Dd	66	kdpqdekilypefnklkenrgylktllslsgwnfgvrtfmlstfsnerfyssviaall	125
OY	69	TEWMDET-LYOENGRKKMPKLTLLALIGMWFCTOKFDMAATANNROTFFVNSAARFL	127
Dd	126	rthgfagdlafflypglrgspardwrfvflllelllqaifneaqtlmrprlllsaaayg	185
OY	128	KRTSPGLDLDWYTPSSQGSFAVDKERFTTLVODLANAFQOEAKTSGKERLLLSAAVPAG	187
Dd	186	phvvqkayearellgrlldfasivsdghawekvtghnsplfs-l--pgdpks-sa_yam	240
OY	188	GYVDAgyEVDKLAQNULDPEFNLMAYDPHGSMWENTGNISPLYRKQESGAASLVNDAAV	247
Dd	241	nywrqlgvppeklmglptrygtfrfnllkasqnelragavpaspbpykytkagfiayeylc	300
OY	248	QQMLQGTPASKRLILCMPTGRSFETLASSSDTRVGAPATSGTPGPPTKEGMALAYEVC	307
Dd	301	cfrvrkakvwndqypyaafkjkewwygdalafgykafrfkirehfganmwcllddffr	360
OY	308	SW-KGATKRIIDQAKPYIFFRDNOWGFEDVESFKTVSLTKOXGLGAGMYMALDDLDDFA	366
Dd	361	gyfcgcpfpflvhtlnnl	379
OY	367	GFSCNGSRYP LIOTLRQEL	385
RESULT	8	A53918	#type complete
ENTRY		chitinase (EC 3.2.1.14)	precursor - braconid wasp (Chelonus sp.)
TITLE		#formal_name Chelonus sp.	
ORGANISM		28-Jul-1995	#sequence_revision 28-Jul-1995 #text_change 03-May-1996
ACCESSIONS		A53918	
REFERENCE		A53918	
#authors		Krishnan, A.; Nair, P.N.; Jones, D.	
#journal		J. Biol. Chem. (1994) 269:20971-20976	
#title		Isolation, cloning, and characterization of new chitinase stored in active form in chitin-lined venom reservoir.	
#accession		A53918	
##status		preliminary	
##molecule_type		mRNA	

KEYWORDS	1-483	#label	KRI
SUMMARY	#cross-references	GB:U10422	
	glycosidase; hydrolase		
	#length	483	#molecular-weight
		52013	#checksum
		5938	
Query Match	34.4%;	Score 1185;	DB 12; Length 483;
Best Local Similarity	42.3%;	Pred. No. 7,24e-199;	
Matches	159;	Conservative 100;	Mismatches 100; Indels 17; Gaps 11;
Db	19	aspmhvcyfigawsvyrtgngkfidinglptclthliysfvyngkdvkylpdsdipgn	78
Qy	19	GSAAKLVCYFFNNMAQYRGEARFLPKDLDSLCHLILYAFAGMTNHOI-STTE--WND--ET	75
Db	79	ldgfgfslrkkhpsvklmavagmgagsvpfsgmasdgateafaguvvklfgyqgfd	138
Qy	76	L--YQEFGLKKMPKLTLLATIGMNFQYKFTDMVATINNRQTEVNSAIRLKRISFD	133
Db	139	gfddwepaqrgspadvkmvklckalkafvq--h----d-yllsaavaeptsask	191
Qy	134	GLIDDMEXPGSGSPAVDKERFTTLVGDLANNAFQGEAQSGKERLLLSAIVPAGQRYVDA	193
Db	192	sydiiaemsgyldfinlmtydthgprwdhtcmhappssahdsngelnklmvaavkylvlgn	251
Qy	194	GEYDKRAQNDLFFVNLAWAFPHGSMWEVTEGNSPLYKROESGAALINDAAVVOOLQK	253
Db	252	gvpkelivvgypaygskftlspnbnkglgavpsagatgagdytenglligyneicemqkag	311
Qy	254	GTPASKLLKMGPTVGRSTTLAASSDTRVGAAPATSGTPPEFTREGMLATYECSW--KGA	312
Db	312	dwevvadnehkvpvaykxngvwsfddlaalakagqfllqgeglsganmwsletddfkgl--c	370
Qy	313	TKQRIQD-QK-VPRIFRDMQWVGSDVDSEFRTKVSYLKQKGLGAGAMWALDLDDPAFGSC	370
Db	371	ge-kyrvkalnsvlg	385
Qy	371	NGGRPLLTQLTRQELIS	386
RESULT	9		
ENTRY	A38221	#type complete	
TITLE	chitinase (EC 3.2.1.14) MFI - nematode (Brugia malayi)		
ORGANISM	Formal_name Brugia malayi		
DATE	04-Nov-1993 #sequence revision 18-Nov-1994 #text_change		
ACCESSIONS	A38221		
REFERENCE	A38221		
#authors	Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Plessens, W.F.;		
	Petler, F.B.		
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552		
#title	Transmission-blocking antibodies recognize microfilarial		
	chitinase in brugian lymphatic filariasis.		
#cross-references	MUID:92179220		
#accession	A38221		
##status	preliminary; not compared with conceptual translation		
##molecule_type	nucleic acid; protein		
##residues	1-504 #label FWH		
##cross-references	NCBI:P85345		
##note	sequence extracted from NCBI backbone		
KEYWORDS	glycosidase; hydrolase		
SUMMARY	#length 504 #molecular-weight 55971 #checksum 7343		
Query Match	31.5%;	Score 1086;	DB 12; Length 504;
Best Local Similarity	42.8%;	Pred. No. 1,85e-179;	
Matches	158;	Conservative 94;	Mismatches 102; Indels 15; Gaps 11;
Db	27	cyttmawgyrdgegfklpignlpglcthllyafakvdelqdsxpfewndetewskmys	86
Qy	26	CYFTMMAQYRGEARFLPKDLDSLCHLILYAFAGMTNHOI-STP--EWND--T-----LYQ	78
Db	87	avtklirenpqllkylsvgygyfsafstlgtaksagqterfikaasafllkmfdgfdl	146
Qy	79	FNNGLKKNPKLKITLLATIGMNFQYKFTDMVATINNRQTEVNSAIRLKRISFDGLD	138

D	b	147	weyr-v-gv-a-e-h-akivreamktatveeaktsqgkrllltaavasgkltigsnve	200
O	y	139	weipssqsppadkcrffttlvodlanaaOQEAQTSVSGKRELLLSAIVPAGQTYYAGEVD	198
D	b	201	sighafdlflmsyblhsweknvdibghklptkgewegsigiftfeafadywaaskmpke	260
O	y	199	KIAQLDLVDVNLMAYHFHSHWSEKYVGHNSPLTKRQEESAAASLNVDAAVQOMLOKGFPA	258
D	b	261	klllgilpmayagwclndpsetaligaasrpsasaktlnpagglaasyelckylkegketv	320
O	y	259	KLIILMPYGRSFTLASSSDTRVGAPATGSGCPGFTEGMALAYEVCSS-KCATIQR	317
D	b	321	hgwygvaymvygdgydygneettirikmkwlkehvggaafiwaldfdgfsgscgkyyp	380
O	y	318	ODAK-PYIFRNOMWGVDVESFETKTYSLKOKIGCFAMWALDLDFAFGSCNOGRYP	376
D	b	381	lmaissel 389	
O	y	377	LQITROEL 385	
R	E	S	RESULT 10	
E	N	T	ENTRY A56596 #type complete	
T	I	T	chitinase (EC 3.2.1.14) - tobacco hornworm	
O	R	G	#formal_name Manduca sexta #common_name tobacco hornworm	
A	N	I	DATE 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 03-May-1996	
A	C	C	ACCESSIONS A56596	
R	E	F	REFERENCE A56596	
A	U	T	#authors Kramer, K.J.; Corpuz, L.; Chol, H.K.; Muthukrishnan, S.	
J	#	j	#journal Insect Biochem. Mol. Biol. (1993) 23:691-701	
#	t	i	#title Sequence of a cDNA and expression of the gene encoding epidermal and gut chitlinases of Manduca sexta.	
#	c	r	#cross-references MUID:93357793	
#	a	c	#accession A56596	
#	s	t	#status preliminary	
#	m	e	#molecule_type mRNA	
#	r	e	#residues 1-554 #label KRA	
#	c	r	#cross-references NCBIIN:136417; NCBIP:136418	
#	e	x	#experimental_source larvae	
#	n	o	#note sequence extracted from NCBI backbone	
K	E	Y	KEYWORDS glycosidase; hydrolase	
S	M	A	SUMMARY #length 554 #molecular_weight 62203 #checksum 4659	
Q	U	E	Query Match 29.9%; Score 1031; DB 12; Length 554;	
B	E	S	Best Local Similarity 38.8%; Pred. No. 1,04e-168;	
M	a	t	Matches 162; Conservative 105; Mismatches 126; Indels 25; Gaps 19.	
D	b	17	vgsdetravcyfsnwayryprgyrgtldipvekcthllysfifgeegnsevlldpel	76
O	y	16	IPMGSAAGTCYFTMAAQYROGEARFLPKDDPSLCITHLIAFAFGMN-H-Q-LST-Tew	71
D	b	77	dvdkgfnfcslshsbvskfmaavggwaegsskyslmvaqsktmsfirsvsfilky	136
O	y	72	N-DETLYDEFGGLKMKNPKLTLTLAGWMNGTGCTFDMAVTANNROTENVSAIRPRKY	130
D	b	137	dfddgidlwepgaadsgsfkdxfilylvgelrrf---lir-qgwel-taavplan	191
O	y	131	SFDGIDLDMERPGS--QSSPAVDKERFTLVQDLANAFQDAQSCKERILLASAIVPAQ	188
D	b	192	fimegyhyvelcgelidalhmsydlgrnwagfadvbaprlxphdhgwayeklnvndglh	251
O	y	189	TYVDYGEVDRKIAQLNDEVNLMAADFHGSWEKVGHNSPLTKRQEESGAASLNVDAVQ	248
D	b	252	lweeqcepnklvnglpbyrgstlsaganmyglgtflinkaagsgdpapytnatgfayy	311
O	y	249	OWLGQGTASKLILIMPTGYRSFTLASSSD-RVGA--P-ATGSGTPGPFRKEGGMALYY	304
D	b	312	eictevdkdsqskvwdegkcpaygkytcwvgyedprsvelmynvllkxyglyantka	371
O	y	305	EVCS--KWG-A--IKQRIQDQKVPFIIFPDNDNWGVDFVESKRTKTVSLIKQKGLGANMYA	359
D	b	372	ldmdfdggl-cge-knpjlikllhkms-sylvpphtenttpdepawarpsptsdpae	426

[illegible]

[illegible]

Qy	78	QEFNLKTKMNRKLTLLAIGMNGTQKTFIDMVAITANNROTFPVNSAIRFLRKISFPGDL	137
Db	170	dweyp-addqetmrvlllkeirsqldayaayq-apy-yhflislaapegehy-sflhm	225
Qy	138	DMEYSGSGSPAVDKERFTTLVODLANAFQOEAQTSKERLLISAAPVAGQTVYDACEY	197
Db	226	sdlsqvldvynlmaydyagswsjsgndanaflampsnpsp-yntqakdylykgvpa	284
Qy	198	DKIAONLDFVNLMADEFSGWEKVTGHNSPLYKRQESGAASLNDAAVOQWLQKQTPA	257
Db	285	sktviqmplygraf-estgg-1gqytsqygs-gsw-englwdy-kvlpkagatqyd	336
Qy	258	SKLLIGMTYGRSFTLASSSDTRRGARPGATSGTGPCTPTEKGMATYEVCSMKGATQ--	315
Db	337	svagayysdpskselisdftpdmlntkvsylknlqigsfmfw	379
Qy	316	RIQDQKVPYIFRDNNQWGFDDVESFKTKVSYLKKGLGAMVW	358
RESULT	14		
ENTRY		QJ01975	#type complete
TITLE		chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus (Aphanocladium album)	
ORGANISM		#formalname Aphanocladium album	
DATE		03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 03-May-1996	
ACCESSIONS		QJ01975	
REFERENCE		QJ01975	
#authors		Blaiseau, P.L.; Lafay, J.F.	
#journal		Gene (1992) 120:243-248	
#title		Primary structure of a chitinase-encoding gene (chl1) from the filamentous fungus Aphanocladium album: similarity to bacterial chitinases.	
#accession		QJ01975	
##molecule_type		mRNA	
##residues		1-423 ##label B1A	
##cross-references		GB:X6104	
COMMENT		This enzyme is essential for the degradation of insoluble chitin.	
GENETICS			
#gene		chl1	
#introns		47/2; 80/2; 97/1	
CLASSIFICATION		#superfamily Streptomyces chitinase chl40	
KEYWORDS		glycosidase; hydrolase	
FEATURE			
1-34			
35-423		#domain signal sequence #status predicted #label SIG\	
SUMMARY		#length 423 #molecular_weight 46072 #checksum 8802	
Query Match		16.7%; Score 576; DB 5; Length 423;	
Best Local Similarity		36.2%; Pred. No. 3,26e-8;	
Matches		101; Conservative 65; Mismatches 95; Indels 18; Gaps 16;	
Db	115	lkkgrmmkwnlslsgytwst-npraassatktktfagsvfgfmkwdgfdgidideyp	173
Qy	83	LKKMPKRLKTLTLAGWFGQKFTDMVATVNNQRTFVNSAIRLRYKSPDGLDMEYR	142
Db	174	-adatcaqumvlllgavseldsyaayq-aky-bhflislaapegnykikfa-e-lg	228
Qy	143	GSGGSPAYDKERFTTLVODLANAFQOEAQTSKERLLISAAPVPG-QTYDAGYEVDKIA	201
Db	229	kvldynlmaydyagswsyctghdanlyanpnpnatp-yntcdavaylingvpankiv	287
Qy	202	QNLDFVNLMAYDHFHSWEKVTGHNSPLYKRQESGAASLNDAAVOQWLQKGPASKLI	261
Db	288	lqmplygsf-qc-eg-1gkpynglgs-gsw-englwdykalpk-agatvkcdttak	339
Qy	263	LQMPYGRSFTLASSSDTRRGARPGATSGTGPCTPTEKGMATYEVCSMKGATQKRIODK	321
Db	340	gcysydstkclisfdtpamistvswlvkyqslgsmfw	378
Qy	322	VYIIF-RD-NQWGFDDVESFKTKVSYLKKGLGAMVW	358

RESULT 15

ENTRY S51369 #type complete  
TITLE Chitinase - fungus (Trichoderma harzianum)  
ORGANISM #formal\_name Trichoderma harzianum  
DATE 15-Jul-1995 #sequence\_revision 19-Apr-1996 #text\_change 03-May-1996

ACCESSIONS S51369  
REFERENCE S51369  
#authors Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell, A.; Pintor-Toro, J.A.  
#journal Curr. Genet. (1994) 27:83-89  
#title Cloning and characterization of a chitinase (CHIT42) cDNA from the mycoparasitic fungus Trichoderma harzianum.

#accession S51369  
#status preliminary  
#molecule\_type DNA  
#residues 1-423 #label GAR

SUMMARY #length 423 #molecular\_weight 46056 #checksum 2000

Query Match 16.3%; Score 561; DB 12; Length 423;

Best Local Similarity 35.7%; Pred. No. 2,14e-78; Mismatches 94; Indels 17; Gaps 13;

Matches 101; Conservative 71; Mismatches 94; Indels 17; Gaps 13;

Db 111 qlfk-vkkanrglkvlisigwtst-nfpaastdanrkfaktaltfmkdwgfdgidi 168  
QY 78 QEFNGLKMKPKLKTLLAIGMNGTOKFTDMVATANNRQTFVNSAIRFLKKSFDGLDL 137  
Db 169 dweyp-adatqasnmlllkevrsqrdayaay-apg-yhflitlaapagkdy-sk1r1 224  
QY 138 DWEXPGSGSPAVDKERFTLVODLANAFQEOQTSGERLLLSAAVPAGQTYVDAGYEV 197  
Db 225 adlqgvldylnlmaydyagsfsltyghdaanfnpsnpatcp-fntdsavkdylingvpa 283  
QY 198 DKIAQNLDFFVNLMAVDHGSWEKVTGHNSPLYKRQESSGAASLNDAAVOQWLQKSTPA 257  
Db 284 nk1v1gmp1ygrsf--gnlaq--lqgtyngygs-gsw--eaagiwdykalkp-agatvgyd 335  
QY 258 SKLILGMPYGRSFTLASSTRTGAPATGSGTGPFTKEGMLAYEVCWKGATKORI 317  
Db 336 svakyyysynsatkelisfdtpdminkvaylksl1gysmfw 378  
QY 318 QDQKVPYIFRD--NQWVGFDVDESFKTKVSYLKQKGLGAAVM 358

Search completed: Fri Jun 27 16:51:51 1997  
Job time : 159 secs.







[illegible]

ID	RESULT	2	STANDARD;	PRT;	504 AA.
AC	P29030;				
DT	01-DEC-1992	(REL. 24, CREATED)			
DT	01-DEC-1992	(REL. 24, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995	(REL. 31, LAST ANNOTATION UPDATE)			
DE	ENDOCHITINASE	PRECURSOR (EC 3.2.1.14) (MFI ANTIGEN).			
OS	BURGI MALAYI.				
CC	EUAROTIA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.				
RN	[1]				
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.				
RP	MEDLINE; 921179220.				
RA	FURHRAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;				
RL	PROC. NATL. ACAD. SCI. U.S.A. 89:11548-11553(1992).				
CC	-1- FUNCTION: THE MFI ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY				
CC	FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-				
CC	FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND				
CC	TRANSMISSION.				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF				
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.				
CC	-1- PTM: O-GLYCOSYLATED.				
CC	-1- KNOWN TO BIND CALCIUM.				
CC	-1- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MFI ANTIGEN CORRESPOND				
CC	WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.				
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
CC	HYDROLASES).				
DR	EMBL; M73689; G156064; .				
DR	PIR; A38221; A38221.				
DR	PROSITE; PS01095; CHITINASE_18.				
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;				
KW	REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.				
FT	SIGNAL	1	22		
FT	CHAIN	23	504	ENDOCHITINASE.	
FT	DOMAIN	23	400	CATALYTIC.	
FT	DOMAIN	401	450	SER/THR-RICH (LINKER).	
FT	DOMAIN	407	448	3 X 14 AA APPROXIMATE TANDEM REPEATS.	
FT	ACT_SITE	148	148	PROTON DONOR (BY SIMILARITY).	
QO	SEQUENCE	504 AA;	55971 MW;	4DAV5EB CRC32;	

[illegible]

Dd	87	avtkirenpbplkvllsygxygnffsaiftgylakbaqctefifikaafifirknmfdgdld	166
Oy	79	PENGKAKNPNKLITLLAIGGNNFGTQFTDVMATANNRQTFFVNSAIRFLRKYSFSDGLDLO	188
Dd	147	weyP-v-gv-a-e-eh-akIveamktafveeaktsqkgri  l taavsagkytldgsynve	200
Oy	139	WEYPSOSCPWADKERFETTLVODLANAFQOEAQRSGERELLSAAVPAGOTYYDAGYED	198
Dd	201	sighnfallflmsydvlbswskenvdlbgklhpctgsvsgidfifntefaadyasaskgmpe	260
Oy	199	KIAQLNDLVNLMAVDHFHSWEXYTGHNSPLYKRDEESGAASLNVDAAYOOMLOKGPAS	258
Dd	261	kiljsipmyagqwtldnpsetaigaasrpsasaaktlpaggtaasyelcikylyegkyte	320
Oy	259	KLIIDMPYGRSFTLASSSDRYCAPAPAGSCOTPOPFKEGMALYYEVCW-KCATQRI	317
Dd	321	bbegvgaymvykgdqwygduaeetirikmkwlkxgygaafiwalddfdcfckscgkypP	380
Oy	318	QDOXK-PFIIFDNQNWGVGGDDVESFKTKSYLKQKGLCGAMWALDLDDPAGFSNCGRYP	376
Dd	381	llnaaisseel 389	
Oy	377	llQTRLROEL 385	

RESULT	3			
ID	CHIT_MANSE	STANDARD:	PR1:	554 AA.
AC	P36362:			
DT	01-JUN-1994	(REL. 29, CREATED)		
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)		
DT	01-FEB-1995	(REL. 31, LAST ANNOTATION UPDATE)		
DE	ENOCCHITINASE	PRECURSOR (EC 3.2.1.14).		
OS	MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HORMWORM).			
OC	EUDAROTA: METAZOA: ARTHROPODA: INSECTA: LEPIDOPTERA.			
RM	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 93357793.			
RA	KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;			
RL	INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).			
CC	-1- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING PROCESS.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-1- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0, BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER LEVELS SEEN ON DAYS 0, 7 AND 8.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS AND GUT.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL HYDROLASES).			
DR	EMBL: U02270: G406049; -;			
DR	EMBL: S64757; -; NOT_ANNOTATED_CDS.			
DR	HSSP: P07254; 1CYN.			
DR	PROSITE: PS01095; CHITINASE.18.			
KM	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	554	ENDOCHITINASE.
FT	DOMAIN	396	453	SER/THR-RICH.
FT	ACT_SITE	146	146	PROTON DONOR (BY SIMILARITY).
FT	CARBOHD	85	85	POTENTIAL.
FT	CARBOHD	303	303	POTENTIAL.
FT	CARBOHD	407	407	POTENTIAL.
FT	CARBOHD	545	545	POTENTIAL.
SO	SEQUENCE	554 AA;	62203 MW;	FA8/F8AD CRC32;

QY 16 IPWMSAAKLVCFYTNMAQYRQGEARFLPKDLDPSLCTHLIYAFAGMTN-H-Q-LST-TW 71  
 Db 77 dvkngrfntfsirspbskfmvayvgaaegskshwaaqstmsfirsavfllkky 136  
 QY 72 N-BETLVEENGKAKMKNPKRTLLAIGMNFQKFTDMVATANNRQTFVNSAIRFLRKX 130  
 Db 137 dfgldldwvpaadaggsfsdskfilylvgeletraf---lrv-gkwel-taavplan 191  
 QY 131 SFGGLDLDWERYPS--QGSFVNDKERTTLVQDLANFQOEAOISGKERLLLSAAYVAG 188  
 Db 192 firmeghyhpelcqealdahvmsydlirngwafadvhsplyktrphdqwayeklnvndglh 251  
 QY 189 TYVDAGYEVQKINQNDVFLMAYDFHGSWEKVTGHNSPILYKQEEGSAASLVNDAVQ 248  
 Db 252 lweekgpcsnklyvgrfygrstfclsegannnylgfclnkeagggdparpynatqfayy 311  
 QY 249 QWLQKGPASKLILGMPTYSRFTLASSSDT-RVGA--P-ATGSGTPGPTKEGMLAYX 304  
 Db 312 elctevdkdsgvttkxwdegskcgyaykfgvwygvedprsvelkmmwllkqylygamta 371  
 QY 305 EVCS--WKG-A--TKQRIQOKVPIYIRDNQWVGFDVESFKTKVSYLKQKGLGAAVMA 359  
 Db 372 ldmddfgql-cge-knpklkllkhms-sylvpphntentpewarppstpsdpe 426  
 QY 360 LDLDFAFGFCNGGRVPLIOTLRQELSLPY-LPSGPELVPRKPGQSPDEHPSPEQ 416  
 RESULT 4  
 ID CH11-BACCI STANDARD: PRT: 699 AA.  
 AC P20533;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1991 (REL. 31, LAST ANNOTATION UPDATE)  
 DE CHITINASE A1 PRECURSOR (EC 3.2.1.14).  
 GN CH11.  
 OS BACILLUS CIRCULANS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WL-12;  
 RA MEDLINE: 90368776.  
 RA WATANABE T., SUZUKI K., OYANAGI W., OHNISHI K., TANAKA H.;  
 RL J. BIOL. CHEM. 265:15659-15665(1990).  
 RN [2]  
 RP MUTAGENESIS.  
 RC STRAIN-WL-12;  
 RA MEDLINE: 93366760.  
 RA WATANABE T., KOHORI K., MIYASHITA K., FUJII T., SAKAI H.,  
 RA UCHIDA M., TANAKA H.;  
 RL J. BIOL. CHEM. 268:18567-18572(1993).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
 CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 CC HYDROLASES).  
 DR EMBL: X57601; G142688; -.  
 DR PIR: A38368; A38368.  
 DR HSSP: P07254; 1CTN.  
 DR PROSITE: PS01095; CHITINASE\_18.  
 KM HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; REPEAT.  
 FT SIGNAL 1 41  
 FT CHAIN 42 699  
 FT DOMAIN 42 460  
 FT DOMAIN 465 549  
 FT ACT\_SITE 560 644  
 FT MUTAEN 200 204  
 FT MUTAEN 200 204  
 FT MUTAEN 200 204  
 FT MUTAEN 204 204  
 SQ SEQUENCE 699 AA; 73677 MW; 5A1777CC CRC32;  
 Query Match 20.3%; Score 698; DB 2; Length 699;  
 Best Local Similarity 40.1%; Pred. No. 3.91e-132;

Matches 119; Conservative 73; Mismatches 85; Indels 20; Gaps 18;  
 Db 144 qlnklyqtpnplktlsvsgtw-snrfsdvaataetrefansavdlkkyndgvdld 202  
 QY 79 EFNGLKMNPKLTLAIGMNFQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLD 138  
 Db 203 wepvrsgldgnskrpedknylllskirekidaagavgk-kyllttsagasya-a 260  
 QY 139 WEPR--GS-QG-S-PADVKEFTTLVQDLANFQOEAOISGKERLLLSAAYVAG 193  
 Db 261 ntelakiaavdwlnmtcydfngawqklsahnaplndypaasaagvpdantfnvaagq 320  
 QY 194 GYEVDRKAQNLVDVNLMAVDFHGSWEKVTGHNSP-L-YKQE-ESGA-A-SLVNDAVQ 249  
 Db 321 hldayvpaaklyvgrfygrwd-gcaagagn-ggyqcltcygsavgtweagsfidylean 378  
 QY 250 WLQKGPASKLILGMPTYSRFTLASSSDTRVGAAPATGSGTPGPTKEGMLAYEV-CS 308  
 Db 379 ylnkngytrvndctavpylynasnkrflsyddaesvgyktaylkakgylgamfwe 435  
 QY 309 W-K-GATKORIDQKVPYIFRD-NGW-VGFDVESFKTKVSYLKQKGLGAAVMA 360  
 RESULT 5  
 ID CH11-APHAL STANDARD: PRT: 423 AA.  
 AC P32470;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE CHITINASE 1 PRECURSOR (EC 3.2.1.14).  
 GN CH11.  
 OS APHANOCADIUM ALBUM.  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ETHM 483;  
 RX MEDLINE: 93013040.  
 RA BLAISEAU P.-L., LAPAY J.-F.;  
 RL GENE 120:243-248(1992).  
 RN [2]  
 RP SEQUENCE OF 35-57.  
 RX MEDLINE: 92136437.  
 RA BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRXGOO Y.;  
 RL CURR. GENET. 21:61-66(1992).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
 CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 CC HYDROLASES).  
 DR EMBL: X64104; G429026; -.  
 DR PIR: J01975; J01975.  
 DR HSSP: P07254; 1CTN.  
 DR PROSITE: PS01095; CHITINASE\_18.  
 KM HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.  
 FT SIGNAL 1 22  
 FT PROPEP 23 34  
 FT CHAIN 35 423  
 FT ACT\_SITE 171 171  
 SQ SEQUENCE 423 AA; 46058 MW; 88456BEE CRC32;  
 Query Match 16.7%; Score 576; DB 2; Length 423;  
 Best Local Similarity 36.2%; Pred. No. 7.00e-103;  
 Matches 101; Conservative 65; Mismatches 95; Indels 18; Gaps 16;  
 Db 115 lkkqnmvmvmlsgvstst-nfpaaasaatrktfagavgfmkdwgfdgldlwepr 173  
 QY 83 LKMANPKLTLAIGMNFQKFTDMVATANNRQTFVNSAIRFLRKYSFDGLDWEPR 142  
 Db 174 -adatqamvlllqavrseldyaagy-akg-htellstaapagdnynklfa-e-1g 228  
 QY 143 GSGGSPAVDKERTTLVQDLANFQOEAOISGKERLLLSAAYVAG-QTYVDAGYEVDKTA 201  
 Db 229 kvldylnlmaydyagswnygtghdanlyanpqnnaatp-yntdavgaylngvpankiv 287

```

OY 202 QNLDVFNLMAYDFHSGMEKVTGHNSPLYKROESSGAASLVNDAVQOVLQGTSPASKLI 261
DB 288 lmplygrsfq-qt-eg-igkpyngigs-gsw--englwdykalpk-agatvcdtdak 339
OY 262 lmplygrsfq-qt-eg-igkpyngigs-gsw--englwdykalpk-agatvcdtdak 321
DB 340 gcyysydpstckeljsfdtpamistkswlksgkglgfmfw 378
OY 322 VYIIF-RD-NQWGFDDVESFKTKVSYLKOKGLGAMVW 358

```

```

RESULT 6
ID CH14-TR1HA STANDARD; PRT; 423 AA.
AC P48827;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE 42 KD ENDOCHITININASE PRECURSOR (EC 3.2.1.14).
GN CHIT42.
OS TRICHODERMA HARZIANUM.
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.
RX MEDLINE; 95269313.
RA GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LLOBELL A.,
PINTOR-TORO J.A.;
RL CURR. GENET. 27:83-89(1994).
CC -1- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH, CELL DIVISION
AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTI-FUNGAL
AGENT.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE
REPESED.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
HYDROLASES).
CC EMBL; 578423; G999376; -.
DR HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOMEN;
KW CHITIN-BINDING.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 34
FT CHAIN 35 423 42 KD ENDOCHITININASE.
FT ACT_SITE 171 171 PROTON DONOR (BY SIMILARITY).
FT CARBOHD 218 218 POTENTIAL.
FT SEQUENCE 423 AA; 46056 MW; 899DA50A CRC32;

```

Query Match 16.3%; Score 561; DB 2; Length 423;

Best Local Similarity 35.7%; Pred. No. 2,55e-99;

```

Matches 101; Conservative 71; Mismatches 94; Indels 17; Gaps 13;
DB 111 qlfk-vkkanrgklvllsfgvwtst-nfpsaastanankfaktitfmkdgfgidid 168
OY 78 QENGLKKNMPKLTLLAIGWNGFTQKFTDMVATANNROTFFNSAIRLKRKTSFGLD 137
DB 169 dweyp-adatqasnmllllkveysqdayaagy-aby-yhllitlaapagkdy-skil 224
OY 138 DMEFPGSGSPAVDKERFTTLVODLANAFQOEQOTSGKEKLLSAVPAQGVYVDAGYEV 197
DB 225 adlgyldylnlmaydagsfslptghdanlfnpnsnpatp-fncdsavkdylngvpa 283
OY 198 DRIQWLDVFNLMAYDFHSGMEKVTGHNSPLYKROESSGAASLVNDAVQOVLQGTSP 257
DB 284 nklvlgmplygrsf--qntag--igqlyngvgs-gsw--eagldykalpk-agatvyd 335
OY 258 SKILLGMPYGRSFTLASSSDTRVGAAPATGSGPGFTREGGLAYEVCMSKMGATKRI 317
DB 336 svakgyysnatckeljsfdtpamistkswlksgkglgfmfw 378
OY 318 QOKVPIYIFRD--NQWGFDDVESFKTKVSYLKOKGLGAMVW 358

```

```

RESULT 7
ID CH11-COCIM STANDARD; PRT; 427 AA.
AC P54196;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ENDOCHITININASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)
DE (CF-ANTIGEN) (CF-AG).
GN CTS1.
OS COCCIDIODES IMMITIS.
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-C735;
RX MEDLINE; 96144270.
RA PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL GENE 167:173-177(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SILVEIRA;
RA YANG C., ZHU Y., MAGEE D.M., COX R.A.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF
N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
HYDROLASES).
CC EMBL; I41663; G1200190; -.
DR EMBL; U51271; G1256769; -.
DR EMBL; U33265; G1255728; -.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW GLYCOPROTEIN.
FT SIGNAL 1 2 POTENTIAL.
FT CHAIN 2 427 ENDOCHITININASE 1.
FT CARBOHD 387 387 POTENTIAL.
FT CONFLICT 15 47 RMLSLRCLCELGARFMFTLSVPAVTVTDIQ -> VOAS
SMSSMPNYPPVPADEGGRFVSRYVFWW (IN REF. 2).
FT CONFLICT 199 199 K -> N (IN REF. 2).
FT SEQUENCE 427 AA; 47629 MW; 6CB9AF73 CRC32;

```

Query Match 15.9%; Score 546; DB 2; Length 427;

Best Local Similarity 34.1%; Pred. No. 9,12e-96;

```

Matches 95; Conservative 71; Mismatches 94; Indels 19; Gaps 17;
DB 115 lknknrlkllsfgvwtv-spnftkfsteggrkfkfadtisklmdgldgfdidweyp 173
OY 83 LKKNMPKLTLLAIGWNGFTQKFTDMVATANNROTFFNSAIRLKRKTSFGLDWEYR 142
DB 174 ede-kqandfvlllkacreaadaysak-hpnqk-kfllitlaapagpny-nk-1klaemd 228
OY 143 GSGSPAVDKERFTTLVODLANAFQOEQOTSGKEKLLSAVPAQ-QTYVDAGYEVDKIA 201
DB 229 klyldfnlmaydfsgvskvsgshmsnfvfsttkp-estpfdsakvdyklyagvpankly 287
OY 202 QNLDVFNLMAYDFHSGMEKVTGHNSPLYKROESSGAASLVNDAVQOVLQGTSPASKLI 261
DB 288 lmplygrsf--ast-dg-igtsfngvgg-gsw--engvdykdmpp-qgqavtledia 339
OY 262 LMPYGRSFTLASSSDTRVGAAPATGSGPGFTREGGLAYEVCMSKMGATKRIODOK 321
DB 340 asysydknkryllsdytklaagkaeyltknmggqm-w 377
OY 322 VYIIF-RDQW-VGFDDVESFKTKVSYLKOKGLGAMVW 358

```

```

RESULT 8
ID CH1A-SERMA STANDARD; PRT; 563 AA.
AC P07254;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE CHITINASE A PRECURSOR (EC 3.2.1.14).
GN SERA.
OS SERRATIA MARCESCENS.

```

Query Match	12.7%	Score 437	DB 2	Length 563
Best Local Similarity	28.8%	Pred. No. 2,88e-70		
Matches	93	Conservative	84	Mismatches 123; Indels 23; Gaps 19
Db	231	pfaaakqkqkytawddpyknfgqldmalbkahpolklpslsgytlscdfffmgdkvix	289	
Qy	57	AFAGNTNHQJSTTEEND--T-LYDFENGLKKNPKLTKLLAIGGWNFGTQKFTDMVATAN	114	
Db	290	-rdrtfvgsvkfegltqwkffdgvdldwefpggkagpnupsgpddgetyylmkelramidq	348	
Qy	115	NKQFVNSAIRFLKRYSP-DGJDDLDWEXRSGGS-PAN---DKERFTTVQDLANAEQO	168	
Db	349	lsaetgr-kyeltsaelsagkdkidvaynvaq-nsmdhifimsydfiygfdlknlgpqt	405	
Qy	169	EAQTSSEKRRLLLSAIVPAGQTYVD-AGEVYENKIAQNIDFVNLIMAYDFHGSWE-KVTGHSN	226	
Db	406	alnawpwpdteayt-tvng-vnallagvgkpgkvavvgamvgigtyngvgnmipftgt	463	
Qy	227	FLYKQDESSGAASLNDVAHQVQMLQKSTPASKLLGLGHPTRGKSTTLASSSDTRVAGPAT	286	
Db	464	atg-pvkgtkwnglndvdyqiaagfimsqwytyldaetaapvfykpsdglittdazavq	522	
Qy	287	GSGRGPPTKEGGMLAYEVCS-WKGAIRKQ-RIDQ-QVPIYR-DN-QWGFDDVESFK	341	
Db	523	akgykvldkqjlgflsweldadn	545	
Qy	342	TKVSYLKQKGLGGAMVWMLDD	364	

RESULT	9	STANDARD:	PRT:	551 AA.
ID	CHIT_NPVAC			
AC	P41684.			
DT	01-NOV-1995 (REL. 32, CREATED)			
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)			
DE	PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).			
OS	AUTOGRAPEA CALIFORNICA NUCLEAR POLYEDROSIS VIRUS (ACMPV).			
OC	VIRIDAE: DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EOBACULOVIRINAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN-C6:			
RA	MEDLINE: 94303173.			
RL	AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERRER M., POSSEE R.D.;			
CC	VIROLOGY 202;586-605(1994).			
CC	-I- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF			
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-I- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES)			
DR	EMBL: L22858; G559195; -.			
DR	PROSITE: PS00014; ER_TARGET.			
DR	PROSITE: PS01095; CHITINASE_18.			
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.			
FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	551	PROBABLE ENDOCHITINASE.
FT	ACG_SITE	305	305	PROTON DONOR (BY SIMILARITY).
FT	CARBOHYD	173	173	POTENTIAL.
FT	CARBOHYD	444	444	POTENTIAL.
FT	SITE	548	551	PREVENT SECRETION FROM ER (POTENTIAL).
SO	SEQUENCE	551 AA;	61368 MM;	DBD124DB CRC32;
Query Match 11.5%; Score 395; DB 2; Length 551;				
Best Local Similarity 29.2%; Pred. No. 1.25e-60;				
Matches 95; Conservative 80; Mismatches 122; Indels 28; Gaps 27;				
Db	221 pwaavqkqkqysawn-p-p-yknfgfqmakiinphlkilpslgygwrlsdp-fyfmhdv	277		
Qy	57 AFAGMTHQSLSTTEWNETTLTY-Q-FENGK-K-KM-NPKIKTLAAGGWNFGTQKTTDWAT	112		
Db	278 -ekrnfvdsvskkfifqvwkffdgvdldwefpgqkganpslgdgdgdaktyllilelram	336		
Qy	113 ANNROTFTVNSAIRFLRKRSF-DGDLDMLEYSGSGS-PAV-D-K-E-R-FTTLVQDLANA	165		
Db	337 lddleagq-grv-yeltsaiaagydkiaavnya-e-aqsklglfllmsydfkgswnldl	392		
Qy	166 FQO-EAQSGKERLLLSAIVPAG-QTYVDAGYEVDKIAQNLDFVNLMAVDFHGSMEKYT-	222		
Db	393 gyqttyva-ps-vmseelytthvvdallkqvgvgnkllvgvanygvgwvgtvntyndy	450		
Qy	223 GMSNLYIKRQESGGAASLNDAAVQOMLQKGTSPASKILLIMPTVIGSFTLASSDTRVG	282		
Db	451 fsgtng-psgsgwedgvydyrdqldklnnyfyfdaaagayfdksqglisfdvs	509		
Qy	283 APATSGSGRPGETKRGGLAYEYVCSMWGAKRQKILDQ-KVPIYF-RDN-QWGEFDVES	339		
Db	510 vlgkvkyvdrnkllgqlfaweidaen	534		
Qy	340 FTKVSYLKQKGLGAGMAYMALDDLD	364		
RESULT	10	STANDARD:	PRT:	820 AA.
ID	CHIT_ALITSO			
AC	P32823.			
DT	01-OCT-1993 (REL. 27, CREATED)			
DT	01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).			
CN	CHIA.			
OC	ALTEROMONAS SP. (STRAIN O-7).			
OC	PROKARYOTA: GRACILICUTES; SCOTOBACTERIA: FACULTATIVELY ANAEROBIC RODS;			
OC	VIBRONACEAE.			
RN	[1]			



Db 425 gsdgqk-idaa-dygeaskyidwvmtcydfifgawak-ngptaphspilaydipqgfn 481  
Qy 184 vppgqrryavagevvdriiaonldfvinlmaydfhsgmewkghnsplykroeesgaaa-sln 242  
Db 482 tadamakfkfskypacklllgfygrgw-gvtgap-ggtatgpat-gty--eagied 536  
Qy 243 vdaavomlqoktgpasklllgmfttgrsftlassdtrvgaprtgsgtpepfkdegmla 302  
Db 537 y-kvlhnsccpatgtlag--tayahcgsnwsydtpatlkskmdaegq1ggaftvefsg 593  
Qy 303 yvevcsmkatqoriodokvpyfrdnomwvgfdvdesfktvsgylkqkglgamwvaldl 362  
Db 594 dtang 598  
Qy 363 ddvag 367

RESULT 13  
ID CHIT\_STRPL STANDARD: PRT: 610 AA.  
AC P11220;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE 63 PRECURSOR (EC 3.2.1.14).  
GN CH7A.  
OS STREPTOMYCES PLICATUS.  
OC PROMAROTIA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92192480.  
RA ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;  
RL GENE 111:69-76(1992).  
RN [2]  
RP SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.  
RX MEDLINE: 88087127.  
RA ROBBINS P.W., ALBRIGHT C., BENFIELD B.;  
RL J. BIOL. CHEM. 263:443-447(1988).  
CC -1- CATALYTIC ACTIVITY: HYDOLYSIS OF THE BETA-1,4 LINKAGES OF  
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- INDUCTION: BY CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE IIT-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
CC (CBD).  
DR EMBL: M82804; G153216; -.  
DR EMBL: M18397; G153209; -.  
DR PIR: A29912; A29912.  
DR PIR: JH0573; JH0573.  
DR HSSP: P07254; ICTN.  
DR PROSITE: P500018; EF\_HAND.  
DR PROSITE: P500561; CBD\_BACTERIAL.  
DR PROSITE: P501095; CHITINASE\_18.  
KM HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.  
FT SIGNAL 1 30  
FT CHAIN 1 610 CHITINASE 63.  
FT DOMAIN 35 140 CELLULOSE-BINDING.  
FT DOMAIN 148 229 FIBRONECTIN TYPE-III.  
FT DOMAIN 236 610 CATALYTIC.  
FT ACT\_SITE 383 383 PROTON DONOR (BY SIMILARITY).  
FT CONFLICT 3 3 F -> I (IN REF. 2).  
SQ SEQUENCE 610 AA: 63974 MW: 2F5E8E35 CRC32;

Query Match 8.3%; Score 287; DB 2; length 610;  
Best Local Similarity 26.4%; Pred. No. 1.37e-36;  
Matches 78; Conservative 63; Mismatches 134; Indels 21; Gaps 16;

Db 320 fnglmlkaeyphklllyfgwtwagg-fpdavknpaafakschdlvedprwadfdq1 378  
Qy 77 ydfebnlktkmnprlktlllaigsnfctokftdmvatanrqrtnvnsairlkrlys-fddgl 135  
Db 379 dldweypancgjlsdcetsapnatsmmkamraefgqdyilta-a---vtadgsdgqk-id 433

Qy 136 dldwvtpgsgoq-s--paydkerfttlviodlanafqoeaotskerlllsaaVPAGOTYVD 192  
Db 434 aa-dygeaskyidwvmtcydfifgawak-ngptaphspilnaydipqgfttadamakfk 491  
Qy 193 AGEVDKIQNLDNFVNLMAVDFHSGMEXKTGHNSPLYKROEESGAAA-SLNDAAVQOVL 251  
Db 492 skypvpadklllgfygrgw-gvtgap-ggtatgpa-gty--eagiedy-kvlnsc 545  
Qy 252 ORGTpASKLLIMPTyGRSFTLASSDTRVGAPATGSGTGPFTKDEGMALAYEVC5MRG 311  
Db 546 patgtvag--tayahcgsnwsydtpatlkskmdaegq1ggaftvefsgtting 599  
Qy 312 atrqriodokvpyfrdnomwvgfdvdesfktvsgylkqkglgamwvaldlldvag 367

RESULT 14  
ID KTXA\_KLULA STANDARD: PRT: 1146 AA.  
AC P09805;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (REF2 PROTEIN)  
DE KUDYEROMYCES LACTIS (YEAST).  
OG PLASMID PGKL-1.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CBS 2359/152;  
RA SOR F., FUKUHARA H.;  
RL CURR. GENET. 9:147-155(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84297209.  
RA STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;  
RL NUCLEIC ACIDS RES. 12:6011-6030(1984).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85037931.  
RA TOKUNAGA M., KAWAMURA A., HISHIMURA F.;  
RL NUCLEIC ACIDS RES. 12:7581-7597(1984).  
RN [4]  
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.  
RX MEDLINE: 87004569.  
RA STARK M.J.R., BOYD A.;  
RL EMBO J. 5:1995-2002(1986).  
RN [5]  
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.  
RX MEDLINE: 90259069.  
RA BRADSHAW H.D. JR.;  
RL NATURE 345:299-299(1990).  
RN [6]  
RP CHITINASE ACTIVITY OF ALPHA-SUBUNIT.  
RX MEDLINE: 91301161.  
RA BUTLER A.R., O'DONNELL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;  
RL EUR. J. BIOCHEM. 199:483-488(1991).  
CC -1- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH  
CC THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE  
CC TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE  
CC TOXIN) TO GAIN ENTRY INTO THE CELL.  
CC -1- PTM: REF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-  
CC SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.  
CC -1- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,  
CC BETA AND GAMMA.  
CC -1- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.  
DR EMBL: X07127; G2829; -.  
DR EMBL: X00762; G2844; -.  
DR EMBL: X01095; G2849; -.  
DR PIR: S07915; S07915.  
DR HSSP: P02877; IHEV.  
DR PROSITE: P500026; CHITIN\_BINDING.





(TM)

**Distribution rights by Intellegenetics, Inc**

t generated

CGACGTCGGACGGCGACTCG..

## Chapter 6

Database 0; Query 0

3333249 seqs, 555961234 bases x 2

Listing first 45 summaries

23:VIR1 24:VIR2

32:BC78 33:BC79

74:PRI8 75:PRI9

94:VRL3 95:VRL4

115:part1 116:part2

mean 11.656; variance 4.533; scale 2.572

ved by analysis of the total score distribution

### Speech Length DB ID

31 1.9 1904 57 AACHILA

## ALIGNMENTS

Human chitotriosidase precursor mRNA, complete cds

## KEYWORDS

ORGANISM

## REFERENCE

**TITLE**

MEDLINE

Submitted (21-JUN-1995) ROLF G. BOOT, DEPAR

15, 1105 AZ, Amsterdam, The Netherlands



RESULT 2 HUMH3G 1741 bp mRNA PRI 24-JAN-1994  
LOCUS Human glycoprotein mRNA, complete cds.  
DEFINITION M80927  
ACCESSION 9348911  
NID glycoprotein.  
KEYWORDS Homo sapiens  
SOURCE Homo sapiens cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1741)  
AUTHORS Hakala,B.E., White,C., and Recklies,A.D.  
TITLE Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family  
JOURNAL J. Biol. Chem. 268 (34), 25803-25810 (1993)  
MEDLINE 94064658  
FEATURES  
source Location/Qualifiers  
1..1741  
/organism="Homo sapiens"  
/cell\_type="chondrocyte"  
/tissue\_type="cartilage"  
1..71  
5'UTR 72..134  
sig\_peptide /codon\_start=1  
CDS 72..1223  
/note="articular 39kDa glycoprotein"  
/codon\_start=1  
/evidence="experimental"  
/product="glycoprotein"  
/db\_xref="pid:9348912"  
/translation="MGVRSQGFVYVLLQCSAYKLYCYTTSMSQYREDSGCFPD  
ALDRFLCHITISFANISNDHIDTEMWDVLYGMLNLTKNPNULKLLSVGWNFG  
SOREKSIANTQSRFTIKSVPEFLRTGFDLAWLYPGRDRQHPFTTLTKEMKAE  
FIKEAPKQKOLLISAALAGKVTIDSSYDLAKISQHDFISIMYDDEGAVRGTTGH  
HSPLEGGEDASPDRESNMVDYAGMLRGAASLVNIGIPFGSGFTLASGETGVA  
PISGPIGRFTKREAGTLYAEICDELGATVHRLLGGOVPAATKGNMVGSDDESV  
KSKVOYLKROLAGMVAALDDPDQSGFCGDDLPFLTMKDALAAT"  
mat\_peptide 135..1220  
/codon\_start=1  
/product="glycoprotein"  
3'UTR 1224..1741  
polya\_signal 1715..1720  
polya\_site 1741  
BASE COUNT 416 a 510 c 447 g 368 t  
ORIGIN  
Query Match 11.0%; Score 182; DB 76; Length 1741;  
Best Local Similarity 63.8%; Pred. No. 1,14e-151;  
Matches 700; Conservative 0; Mismatches 374; Indels 24; Gaps 10;

QY 349 ATATGTTAGCCAGCGCCACACACCGTCAGACTTTTCTCAATCGCCATCAGTTTCTGC 408  
Db 454 gaaccatgctttagatggcctgagacctgctgctgctacccttga-----ggag-- 505  
QY 409 GCAAAATACAGCTTTGACGGCCTTGACTGACTGGAGATACCAGAACCCAGGGAGGCC 468  
Db 506 -----agacaaacagcatttaccacctatacaaggaatatgaagccgaattatata 558  
QY 469 CTGCCCTAGACAGAGCGCCTTACAAACCTGTGTAGAGACTTGGCCAAATGCCCTTCACG 528  
Db 559 aggaagccag--c--caggaagaagcagctctctcagcagcagcagcagcagcagcagcagc 615  
QY 529 AGGAAGCCCAAGCTTCAGGAGGAAGAACGCTTCTTGTAGTTCAGCGGTTCACGCTGGGC 588  
Db 616 aggtacacattgacagcagcattgacattgccaagatatcccaacacctgatttcaata 675  
QY 589 AGACCTATGTGATGCTGTGATGAGAGTGAGAGTGAACAAATCCCGAAGACCTGATTTGTCTCA 648  
Db 676 gcatcatgacctagatttcatatgagcctgctgctgctgctgctgctgctgctgctgctgctg 735  
QY 649 ACCTTATGGCTTACGACCTTCCATGGCTCTTGGAGAGAGTCAAGGATCAACAGGACCCG 708  
Db 736 tgtccgagtgcaagagagatgacagctcctgacagattcagcaacacctgactatgctgtg 795  
QY 709 TCTACAAGAGGCAAGAGAGAGAGTGTGACAGACCGCTCAACGTGATGCTGCTGTGC 768  
Db 796 ggtacatgttgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 855  
QY 769 AACAGTGGCTGAG 828  
Db 856 ggaagagcttaccctctgctcttctc--tgagacty--gttvtgagcccaatcagagc 912  
QY 829 GACGCTCTTACACACTGGCTCTCTCATACAGACAGAGAGTGGGGGCCAGCACAGGCT 888  
Db 913 cgggaattccagagcagctgctcacaagaagcagagcagcagcagcagcagcagcagcagcag 972  
QY 889 CTGGCACTCAAGGCCCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948  
Db 973 acttcctccgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1032  
QY 949 CTTGAGAG-G-GG-GCCACCAACAGAGATCCAGAGATCAGAGAGAGAGAGAGAGAGAGAG 1005  
Db 1033 agggcagcagctggtgtagatagcagcagcagcagcagcagcagcagcagcagcagcagc 1092  
QY 1006 GGGACAAACAGTGGGTGGCTTGTATGATGTGAGAGCTTCAAAACCAAGCTCAGCTATC 1065  
Db 1153 gctcctctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1170  
QY 1126 GCTTCTCTCTGCAACACAGG 1143

RESULT 3  
LOCUS HSD49835 1418 bp mRNA PRI 25-JUL-1996  
DEFINITION Human YKL-39 precursor mRNA, complete cds.  
ACCESSION U49835  
NID 91457940  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1418)  
AUTHORS Hu,B., Trinn,K., Figueira,W.F. and Price,P.A.  
TITLE Isolation and sequence of a novel human chondrocyte protein related to mammalian members of the chitinase protein family  
JOURNAL J. Biol. Chem. 271 (32), 19415-19420 (1996)  
MEDLINE 96325055  
REFERENCE 2 (bases 1 to 1418)  
AUTHORS Price,P.A.



FEATURES	source	Location/Qualifiers
		1..1434
		/organism="Homo sapiens"
		/map="1p13.3"
		/chromosome="1"
		/tissue_type="Scenes placenta"
		/dev_stage="8 to 9 weeks"
		/clone="Est 257753"
5'UTR		1..36
		/gene="HDWTCCHIT"
exon		1..76
		/gene="HDWTCCHIT"
		/note="exon 1a: first alternatively splice exon 1; contains a Kozak consensus initiator methionine"
		/number=1
		37..114
		/gene="HDWTCCHIT"
sig_peptide		37..1209
		/gene="HDWTCCHIT"
		/note="first alternatively spliced form, see GenBank Accession Number U58515 for the second alternatively spliced form"
CDS		/codon_start=1
		/product="chitinase precursor"
		/db_xref="PID:91439566"
		/translation="MGATMDQKSLMAGVVLLLLDGGSAKLYCYFTNMSQDRQEPGKFTENIDPLFCSHLIYSFASIEENKVIIRKSEMYLQITNSLTKNPKILLISGILASGKGFHMKVDSSTSRLEFINSLIFLNHNPDGIDVMTYDPDCKENKFLVLIHGLAFAPOKDFPKMKERLLNAGVSAGROMIDNSYQEKLAKDIDFILNSFDRHGSNKEPILITGNPSLPSKGMDRGPSTYNEIAYGVIIHKMPSEKYVMGLPIYGHSTFLTASAEYTPAGPAGPITESGFLAYEELCOPLKGAKTIRLDQDOVPAAVKNQWVGYDYVKSAMETKVIQPLKMLNLGAMISIDMDFTGKSCNPGPYPLVQAVKRSLSL"
exon		77..106
		/gene="HDWTCCHIT"
		/number=2
exon		107..309
		/gene="HDWTCCHIT"
		/number=3
mat_peptide		115..1206
		/gene="HDWTCCHIT"
		/note="first alternatively spliced form, see GenBank Accession Number U58515 for the second alternatively spliced form"
		/product="chitinase"
		310..365
exon		/gene="HDWTCCHIT"
		/number=4
exon		366..516
		/gene="HDWTCCHIT"
		/number=5
exon		517..643
		/gene="HDWTCCHIT"
		/number=6
exon		644..771
		/gene="HDWTCCHIT"
		/number=7
exon		772..954
		/gene="HDWTCCHIT"
		/number=8
exon		955..1071
		/gene="HDWTCCHIT"
		/number=9
exon		1072..1433
		/gene="HDWTCCHIT"
		/number=10
3'UTR		1210..1433
		/gene="HDWTCCHIT"
polyA_signal		1407..1412
		/gene="HDWTCCHIT"
polyA_site		1434
		/gene="HDWTCCHIT"
		/note="32 A nucleotides"

Query Match	9.5%	Score 157	DB 74	Length 1434
Best Local Similarity	62.9%	Pred. No. 1.33e-125		
Matches 710	Conservative	0	Mismatches 391	Indels 27
Gaps				10
Db	69	ctggcaggtgtagtgcgttcgtctgtcctccagggagatgcctacaactggtttg	128	
Qy	44	ctggcaggttcataggtcctgtgatgatcccatggggctcctctcaaaaactgcttg	103	
Db	129	ctacttctccactggtgtcccgagcccgagaaacccaataatccacctgagat	168	
Qy	104	ctacttccaccactgtggcccgagctacagacaggggagcctgcctccgccaaggcctt	163	
Db	189	tgacccttcctatgctctcatccatcatctatcttcggcagcatcgtaaaacaaga	248	
Qy	164	ggaccctcagccttctgcaccacccatctatgccttgccttgccatgacacacac	223	
Db	249	tatcatcaaggacaagatgaatgtagtcttaccagaaccatacaagctccaaaac	308	
Qy	224	gagcacacactagtgatgtagacgagactctctacacagagttcaatggcctga	283	
Db	309	gaatcccaactgaaatattcctctgtccatctggagggtacctgtttggtccaa	368	
Qy	284	gaatcccaaggtgaaagaccctgtttgacacatcgaggcctggaatttcgac	343	
Db	369	ccaacctatgtgtgattctctctacatcaagcttggaattcaactccataatc	428	
Qy	344	cacacattttgttagccacggccacacacacccgtcagacctttgtcattcggc	403	
Db	429	tctggagacataacttctgtagtgcgtgtagtgaagcttgatctaccaa	485	
Qy	404	tctggcgaatatacagacttgcagggccttgacctgtgactgggagtagccag	463	
Db	486	-agaa-----a-acactcat---ttcaactgctgattcaatgaatgaagaac	533	
Qy	464	gagccctccctgtagacagagacgctcttcacaaacctgtgtagagacttg	523	
Db	534	tcaaggaagattcaacaataccacaaggaaagctctctctgctcgggcgat	593	
Qy	524	ccagcaggaagaccacagacctcagggaaagaaagcccttcttgatgtagcag	583	
Db	594	aggaggacaatgatgatlaaacgctatcaagtgtgagaacttgcagaagat	653	
Qy	584	tgggcagacatgatgtagtgcgtatgcagagctggacaaaatccccaagac	643	
Db	654	catcaacctcctccttcttgaacttccatcagtggtctctggaaaagccct	713	
Qy	644	ttctcaactttatggccttcagacttccatgctcttgaggagag-----	697	
Db	714	caacgacctcttgagcaaggggtgacggagaagggcccaagctcctactaca	773	
Qy	698	taaacgccccctcttacaagagggcaaaagagtggtgcacacacccacac	757	
Db	774	atactcttggtgtagtcatatcaataaggaatgcatacagaaggttgctat	833	
Qy	758	tgctctcttgcaacagtgctgcagaaaggagaccctgcacagacagtgat	817	
Db	834	cccacataatggcactccttcacactgacct--tg--cagaaaccaacgt	890	
Qy	818	gccttaccttaccagaccctcttcaactggcctctctcatcagacacagat	877	
Db	891	tgcccttgcccttgaggtgtgtgacacatacaagagttctcagcttctcg	950	
Qy	878	accacacaggtctgaccttccagcccttccacaaagagagagtgctgacct	937	
Db	951	tgaagcttgccaggttccctgaagagagccaaagatcacgcggtccaggt	1010	
Qy	938	tgaatctgc---tcttggaagggggccacccaacacagaaatccagat	994	
Db	1011	ctaagcagtcacagggnaacagtggtgtggtctatgatgatgtgaagat	1070	
Qy	995	cttacttcttccggcgaacacagtggtggtgctttgatgatgtgagagct	1054	

[illegible]



QY 819 CCTACTGAGGAGCTCTTTCACACTGCGCTCTCATCAGACACGAGTGGGGGCCCA 878

Db 898 gcatcagggccaggaataccagcggtcaccaaggaagaaggaactcttactat 957

QY 879 GCCACAGGGCTTGGCACTCAGAGCCCTTACCCAGGAAGAGGAGTGTGCTACTAT 938

Db 958 gagatctgtagcttcctccaaggaagcagctcgttagaccccttgcagcaggtcccc 1017

QY 939 GAAGCTG---CTCTGGAAAGGGGGCCACCAACAGAGATCCAGATCAGAAAGTGCCCC 995

Db 1018 tatgcacacaagggcaacagcagtgggtggtatgatgaccagagagcgctcaaaaaaag 1077

QY 996 TACATCTTCGGGACACCAAGTGGGTGGCTTGTGATGATGAGAGCTTCAAAACCAAG 1055

Db 1078 gcaaggtacctgaagagcgagcgtggtgctatggtgtgtgaccccttgacttgat 1137

QY 1056 GTCAGCTACTTAAGCAGAGGAGACTGGCGGGCCATGCTGTGGCACTGACTTAAAT 1115

Db 1138 gacttcggggcaactctg 1157

QY 1116 GACTTTGCCGGCTTCTCTG 1135

RESULT 7

LOCUS SSU19900 1733 bp mRNA MAM 02-FEB-1996

DEFINITION Sus scrofa 38kDa heparin-binding glycoprotein mRNA, complete cds.

ACCESSION U19900

KEYWORDS 6643470

SOURCE

ORGANISM

Plg.

Sus scrofa

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Artiodactyla; Suiformes; Suidae; Sus.

1 (sites)

Shackleton, L.M., Mann, D.M. and Mills, A.J.

Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling

J. Biol. Chem. 270 (22), 13076-13083 (1995)

95286589

2 (bases 1 to 1733)

Mills, A.J.T.

Direct Submission

Submitted (17-JAN-1995) Albert J.T. Mills, Biological Sciences, University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA

FEATURES

source

1..1733

/organism="Sus scrofa"

/note="in Lambda Zap II"

/clone="PB538k"

/tissue-type="smooth muscle"

/dev stage="adult"

67..1218

/note="gp38k"

/codon\_start=1

/product="38kDa heparin-binding glycoprotein"

/db\_xref="PID:9643471"

/translation="MGLRAOTGFALVILSCAYKIVCYTSMQREBDSGCFD

ALDPFLCHITISFANISNNEIDTLEMDVLTIDTLNLRNPKLTLSTYSGNNG

SOREKIASNTOSRTFTKSVPELRTGFGDLDAWISPERDRRLHTLVKEMKAE

FREKALPETERLLSGAVSAGVAIDROYDAIQSDHDFISLTYDHGAMRQTG

HSPLRGOGDASDRFSNADYAVSYRLGAPANLWAGIPTFGRSFTLASKIDVGA

PASPGIGREFTEKREGLAYEICDFLOGATVRRPLGGOVPAATGNGMWGDDQESV

KNKAKYLSRDLAGAMVMTLDIDDFRGNFCQNLRFPLSAIKDVLAA"

BASE COUNT 408 a 507 c 463 g 355 t

ORIGIN

Query Match 9.3%; Score 154; DB 48; Length 1733;

Best Local Similarity 62.5%; Pred. No. 1,69e-122;

Matches 687; Conservative 0; Mismatches 389; Indels 24; Gaps 9;

Db 79 gtggctcagacaggtttgtgcccctgtgtctgtccagagctgtgtcctacaagctg 138

||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||

QY 39 GTGCGCTGGAGCAGTTTCATGCTGCTGTGATGATGCCATGGGGCTCTGTCGAAACTG 98

Db 139 gtttctactataaccagctgtgtctcaagtcaggaggtgtatggaagctgtctccagat 198

QY 99 GTCTCTACTTTCACCAACTGGGCCCAGTACAGACAGGGGGAGGCGCTCTTCTGCCCCAAG 158

Db 199 gccatcgacccctctctcagccacacatcatctcaagcttgcacataaagacaacat 258

QY 159 GACTTGGACCCAGCCTTTGACCCCACTCATCTACGCTTGCTGGCATGACCAACAC 218

Db 259 gagattgcacaccttgagtggaatgatgtgaacgtctatgaacactgaacactcaag 318

QY 219 CAGCTGAGACACACATGAGTGAATACGAGACTCTCTACAGGAGTTCAATGGCTGAAG 278

Db 319 aacaggaaccccaacctgaagaacctcctctgtctgttgaagatggaacttgttctaa 378

QY 279 AAGATGAATCCCAACTGAAGACCTGTATGACATCGGAGGCTGGAATTTACAGACTCAG 338

Db 379 agatttccaaaatagctcccaacatcagagtgcgagacttcatcaagtgcgtgca 438

QY 339 AAGTTCACAGATATGTATGACCGCCCAACACCGTCAACCTTTGTCAACTCGGCCATC 398

Db 439 ccattccctgcgagaccatggtcttltgatgtgactggaactagacctggtcccttg--gc 496

QY 399 AGCTTTCTGCGCAATATACAGCTTTGACGGCCTTGACTGAGAGTACCCAGGAAGC 458

Db 497 -----ggag-----agacaagcgcacatcaccactctagtcaaggagatgaagct 543

QY 459 CAGGGAGCGCTGCCGTAGACAGAGGCGCTTCAACACCTGTGACAGACTTGGCCAT 518

Db 544 gagtttgttaaggagccctg--cct--ggaacagagcgctctctgacgagcggtg 600

QY 519 GCCTTCAGACAGAAAGCCCAAGCCTCAGAGGAAGAACGCTTCTCTGATGTCACAGCTT 578

Db 601 tctgaagggaagtgccattgacagaggtatgacatgcgcccagatatctcaacactg 660

QY 579 CCAGCTGGGCAACCTATGTGATGCTGTGATGAGTGAAGTGAACAAATGCCAGAACTG 638

Db 661 gacttcacagactttagacctatgacttccacggtgtcctggtgcgaactcaagaaac 720

QY 639 GATTTGTCAACCTTATGGCTTACGACTTCACTGCTTGTGGAGAAAGTCAACGGGACAT 698

Db 721 cacagtcacctgtccggagcgaggagtgtaagttccgaagatctagcaaatcagac 780

QY 699 AACACCCCTTAAAGAGGCAAAAGAGTGTGACGACGCCACTCAACCTGCAAT 758

Db 781 tatgtgtgactatgtctgagagctgggggcccgaagcaataagctgtgtatggtatc 840

QY 759 GCTGCTGTCAACAAGTGTGCTGAGAAAGGAGCCCTCGCAGCAAGCTATCTTGCAATG 818

Db 841 cccactttggagagagcttcaacttgcctcttcca--agaag--ac--gtggagaccccg 897

QY 819 CCTACTAGGAGCTCCTTCACTGCTCTCTATCAGACACCAAGATGGGGCCCA 878

Db 898 gcatcagggccaggaataaccagcggttccacaaaggaagaaggaactcttactat 957

QY 879 GCCACAGGGCTTGGCACTCAGAGCCCTTACCCAGGAAGAGGAGTGTGCTACTAT 938

Db 958 gagatctgtagcttcctccaaggaagcagctcgttagaccccttgcagcaggtcccc 1017

QY 939 GAAGCTG---CTCTGGAAAGGGGGCCACCAACAGAGATCCAGATCAGAAAGTGCCCC 995

Db 1018 tatgcacacaagggcaacagcagtgggtggtatgatgaccagagagcgctcaaaaaaag 1077

QY 996 TACATCTTCGGGACACCAAGTGGGTGGCTTGTGATGATGAGAGCTTCAAAACCAAG 1055

Db 1078 gcaaggtacctgaagagcgagcgtggtgctatggtgtgtgaccccttgacttgat 1137

QY 1056 GTCAGCTACTTAAGCAGAGGAGACTGGCGGGCCATGCTGTGGCACTGACTTAAAT 1115

Db 1138 gacttcggggcaactctg 1157

QY 1116 GACTTTGCCGGCTTCTCTG 1135



	RESULT	8	
AC	SSGEP38KD	standard; RNA; MAM: 1733 BP.	
NI	247803:		
DI	9634097		
DT	19-JAN-1995 (Rel. 42, Created)		
DE	28-AUG-1996 (Rel. 49, Last updated, Version 6)		
KV	S.scrofa 38kda heparin-binding glycoprotein.		
OS	Gp38k; heparin-binding glycoprotein.		
OC	Sus scrofa (domestic pig)		
NC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
RA	Theria; Eutheria; Artiodactyla; Suidae; Suidae.		
RP	[2]		
RL	Submitted (19-JAN-1995) to the EMBL/GenBank/DBP databases.		
RT	Millis A. J., University at Albany, Biological Sciences, 1400		
RN	Washington Avenue, Albany, NY, USA, 12222		
RM	[3]		
RX	MEDLINE: 95286589.		
SA	Shackelton L.M., Mann D.M., Millis A.J.;		
RI	"Identification of a 38-kDa heparin-binding glycoprotein (gp38k)		
RT	in differentiating vascular smooth muscle cells as a member of a		
RL	group of proteins associated with tissue remodeling";		
J	J. Biol. Chem. 270:13076-13083(1995).		
FH	Key Location/Qualifiers		
FT	source		
FT	1..1733		
FT	/organism="Sus scrofa"		
FT	/clone="PBS38k"		
FT	/dev_stage="adult"		
FT	/tissue_type="smooth muscle"		
FT	/clone_1lb="lambda zap ii"		
FT	67..1218		
FT	/product="38 kDa heparin-binding glycoprotein"		
FT	/db_xref="PID:g634098"		
FT	/translation="MGELRVAGTGFVALVLQSCAAKYLVCYYTSMSOYREGDGCFFPL		
A	IDPLECTHIISFANISNNEIDTLEMNVDVLTDLTLNLKNPNKLKTLLSVGNWFGS		
O	RFSKIASVTQSRRFIKSVPFLRTHGDDGLMISGRDKRHLLTVLKEMAEFPI		
R	EALPGERLLLSGAVSACKVAIDRGYDIAQISOHLDFISLLTYDFHGAMROTDGHSH		
L	FRGGDASDSRFSNADYAVSYVLRLGAPANKLVNGIPTFGRSFTLASSTKVDPGAPASG		
P	GIPRFTRKEKILAYVELCELDPLQGATVVRPLPGQQVPPVATRKONQWGVYDDSEYNNKKAK		
Y	LKSRQLAGAWMYTLDLDFRNFCGQINLPFLTAIKDVLA**		
SO	Sequence 1733 BP: 408 A; 507 C; 463 G; 355 T; 0 other;		
	Query Match 9..3%; Score 154; DB 15; Length 1733;		
	Best Local Similarity 62.5%; Pred. No. 1,69e-123;		
	Matches 687; Conservative 0; Mismatches 389; Indels 24; Gaps 9;		
D	79 gtggtctcagaacaggttttgttgccctgtgtgcgtgccagaagtgtgtgtcctacaagctg 138		
Q	39 GTGGCGCTGGGGCAGGTTCATGCTCTCCGTGCAIGATGCCATCGTGCGCTCTCCAATAACG 98		
D	139 gttgtctactatcaacagctgtcttaagcacggagggtgatgggaagctgtctccagaat 198		
Q	99 GTCCTCTACTTACCACACTGGCGCCCAAGTACAGACAAGGGGAAGCTTCCTCTGCCAAG 158		
D	199 gccatcgcccccttcctctgcaccacacatcatcaacagctttgtccaacaataaacacaa 258		
Q	159 GACCTTGGACCCCAGCCTTTGACACCCACCATCACTACGCTTCGTGCATGACCAACAC 218		
D	259 gagattgcaccttgtagtggaatatgtgagagctctatagaaaacactgaacacactaaag 318		
Y	219 CAGCTGACACACTGAGGAGAATACGAGACTCTCTACCAAGAGCTTCAATGCGCTGAAG 278		

D	b	319	aacaggaaccccaacttaagaaccttcctcgtctgcttgtagagatgaaagtggtttcca	378
O	y	279	AAGATGAATCCCAAGCTAACAACCCCTGTTAACCATCGAGGCTGGATTTCAGACTCAG	338
D	b	379	agatttccaaaataagctcccaacaactcagatcgcgagacttcatcaagtcggtcca	438
O	y	339	AAGTTCACAGATATATGTATGCACGACCACAAACCGTCAAGACTTGTTCACCTCGGCATC	398
D	b	439	ccattcccgcggagaccataggctttgatgatgtaactagctcgttatccccg--gc	496
O	y	399	AGGTTTGCGCCAATAATACAGCTTTGACGCCCTTGAACCTTGACTGGAGTACCAGGAAGC	458
D	b	497	-----ggaag-----agacacgagatctcaaacactcagtcgaagaagatgaagct	543
O	y	459	CAGSGGACCCCTGCCGTATGACACAGAAGAGCGCTTCACAAACCTGTGTACAGACTTTGGCCAT	518
D	b	544	gagtttgtaaagyaagccctg-cc-t--gnaacagagcygctcctcgtctcagcysagcgtg	600
O	y	519	GCCTTCCAGCAGGAAGAACCCACAGCCTCAGGGAAGGAACGCTTCTTGTGATGACGCGGT	578
D	b	601	tctgagaggaagygltcgcaattgacagagagcctatgacatcgcccaagatatccaaacctg	660
O	y	579	CCAGTGGGCGAGACCTATATGTGAATGCTGGATACAGAGGTGCAAAATTCGCCAACACCTG	638
D	b	661	gacttcacagcagcttctaactatgaacttccagctcagctgctcgtcgccaactacagagac	720
O	y	639	GATTTTGTCAACCTTATGGCTTAGACACTTCATGCGTCTTTTGGAGAAAGTACCGGCAT	698
D	b	721	cacagtcgccctggtccagagccaggyggaigtcaaattccagatlcagcaatgcagac	780
O	y	699	AACACCCCCCTTACAAAGAGCAAGAAAGAGTGGTGACAGCACGACCTCAACGTGAT	758
D	b	761	tatgtctgagactatgctctgagctcgtgggcccacagcaataagctcgtgtatggtatc	840
O	y	759	GCTGCTGTGCAACAAGTGGCTCCAGAAAGGGACCCCTCGCACCAAGCTGATCTTGGCATG	818
D	b	841	cccaatttggagaagagctcacitggtcccccttcca-agaag-ac-gtggagagcccg	897
O	y	819	CCTACTCTAGGAGAGCTCTTACACTGGCTCTCTCATCAGACACCAAGATGGGGCCCA	878
D	b	898	gcatacagggccaaggaataccagagccggttccaccaagayaaggaatoccttgcctatat	957
O	y	879	GCCACAGGGTCTGGGACATCCAGGCCCTTTCACAAAGGAAGAGGATGCTGGCTACTAT	938
D	b	958	gagatcgtgacttcctccaagsgccaagtlccqtagaaccccccttgcagcaggtlcccc	1017
O	y	939	GAAAGCTG---CTCTGGAAAGGGGGCCACCAAAAGAAATCCAGATCAAAAGGTGGCC	995
D	b	1018	tatgscaccaagsgcaaacagttggtggtatgatatgacagggagdgaggttcaaaaaag	1077
O	y	996	TATACTTTCGGGGACAAACGATGGTGGCTTTGATGTGTGGAGAGGTTTAAAAACAG	1055
D	b	1078	gcaaatglactctaagaagcagcagctgctggtctatgctatgctgctgagccctgactgcat	1137
O	y	1056	GTCACCTATCTGAAGCAAGAAGGAGCTGGCGGGGCGANTGRTCTGGCAGCTGGACTTAGAT	1115
D	b	1138	gacttcggggcaactctg 1157	
O	y	1116	GACTTTGCCGGCTTCTCCTG 1135	

RESULT	9	KMBRP39	1616 bp	RNA	ROD	28-NOV-1995
LOCUS		M.musculus mRNA for BRP39 protein.				
DEFINITION		X93035				
ACCESSION		g1085065				
NID		brp39 gene; BRP39 protein.				
KEYWORDS		house mouse.				
SOURCE		Mus musculus				
ORGANISM		Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;				
		Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorphae; Muridae;				
		Murinae; Mus.				

REFERENCE	1 (bases 1 to 1616)
AUTHORS	Morrison, B. W. and Leder, P.
TITLE	neu and ras initiate murine mammary tumors that share genetic markers generally absent in c-myc and int-2-initiated tumors
JOURNAL	Oncogene 9 (12), 3417-3426 (1994)
MEDLINE	95060797
REFERENCE	2 (bases 1 to 1616)
AUTHORS	Morrison, B. W.
TITLE	Direct Submission
JOURNAL	Submitted (14-NOV-1995) B. W. Morrison, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES	Location/Qualifiers

```

/misc_feature
    /organism="Mus musculus"
    /strain="FVB/N"
    /dev_stage="adult"
    /tissue_type="breast"
    /cell_line="SMF murine breast tumor"
    /clone_idb="lambda ZAP II"
    66..128
    /gene="brp39"
    /note="leader sequence"
    66..121
    CDS

```

/note="homology to glycosyl hydrolases"  
 /codon\_start=1  
 /product="BRP39 protein"  
 /db\_xref="PID:e211795"  
 /db\_xref="PID:g1085066"  
 /translation="MGMRAALTGFAVLMILQSCAAYKLCVCFITSMQSYRCVGSFLPT  
 AIDPLICTIHTIYSPANISDDMLSTWENDSDNSDKLTKLTKTRNTNTLTISVGWKKR  
 GEKSPSEIASLTERRTAFVRSVAFPLFSGYGDGLAMLYRLDKQYFSTILKEINL  
 EFTKEVQGREKLLLSALSGKVAIPDGYDIAOIAHLDIEINLTAYDHFHWMOITQ  
 HSPSLFOGQKDRDRXSNVAYVOYMRALGAQSKILMGIPFGKSTLSSNSOIGD  
 APISGEIIPGRTREAGTALAYEICDPIKGAQEVRLTSERKPYATKQMWGYEHKES  
 VKNVGFLEKELKLGANVMALDIDDFQCTCPKEFFPLTAIKDIA

Query Match	8.0%;	Score 133;	DB 81;	Length 1616;
Best Local Similarity	62.4%;	Pred. No. 6.95e-101;		
Matches	673;	Conservative	0;	Mismatches 378;
			Indels	27;
			Gaps	13;

Db	88	caggctctgggtccctcctatgctgtcccaagacgtctctgctgatacaagctgtgtcgtact	147
Qy	49	CAGGTTTCATGTGCTCTGCTGTATGATTCCTCCATGGGGCTTGCTGCAAAACTGATGCTTACT	108
Db	148	tcaccagctgtgtcccaagctaccgggaagcgtgttgaagctcttaccacagccatccaac	207
Qy	109	TCACCAACTGGGGCCAGTACAGAAAGGGGAGAGCGCTCCGCTTCCTCCCAAGACTTGGACC	168
Db	208	cttctctgtgaccacacatcatctacagcgtttgccaatcagcagcgaacaatgttta	267
Qy	169	CGAGCTTTGGACCCACACTCATCTATCAGCGCTTCGCTGGCAGTACCAACAC -CA -GCTGA	225
Db	268	gcaatggtggtatggaatgaacgaagctgcgaactatgcaagcgtgataaacctgaagcagaa	327
Qy	226	GCACCCAGTACGTGAATGATGCAGAGCTCTACCCAGGAGTTCAATGGCTCATAAGACATGA	285
Db	328	acaccaactctaaagaccctctcgtctgttggagggtggaatttggsgaaaaagatttc	387
Qy	286	ATCCCAAGCTAAGACCTGTGTAACCAACGGAGCGTGGAAATTTCAGACTCAGAAAGTTCA	345
Db	388	ccgaagtatgcctcaacaactagaagacgaactgtcttcgtctcgtgtgtagcccccttc	447
Qy	346	CAGATATGATGACCAAGGCCAACAAACCTCAGACCTTTGTCAATCGCCATCATGAGTTTC	405
Db	448	tgcgtctcttatgctttagatgggtcgtgatactcgcctcgtctctaacct- - - - -C- - - - -	495
Qy	406	TGCCCAATATAGAGCTTTGACGGCCTTGACTTGACTGGAGATACCAAGAAAGCCAGGGGA	465
Db	496	gct-taa-g-agaagaagcagattcttcacacctgatacaagaaatgaatgacgaattca	552

QY	466	GCCTGCGGTGAGACAAGAGAGCGTTCCACAACCCGTGATACAGAGACTTGGCCAAATGCGCTTCC	5253
Db	553	caaaaggaggttcacg--c-caaggcagagaagaatcttcgctcagcgcaagcttctgcaag	6099
QY	526	AGCGAAGAGCCCAACACCTCAGAGGAGGAAGAGCCCTTCTTCTTGAGTGCAGCGGTTCACACTG	5855
Db	610	gaaaagtgccatcgacacgtgcgtatgacatgcgcgcagatagaccgaacactgtgattta	6659
QY	586	GGCAGACCTATGTGGATGCTGGATACAGAGTGTGGACAAATTCGCCAGAACCTGGATTTTG	6455
Db	670	tcaatctcatcattgacacgatcttccatgagtlctgcgcacaatacagcagccacacagcc	7299
QY	646	TCAACCTTATGGCTTACGACTTCATGGCTCTTGGAGAAAGTGCACGGGACATATACAGCC	7050
Db	730	ccctcttccaaagccagaaaggaacactaggtttgcagatacagaatgtaactatgpcg	7899
QY	706	CCCTCTTACAAAGAGCAGAAAGATGATGGTGCAGCAGCCAGCCCTCAACGTGATGCTGCTG	7655
Db	790	tgcagtatcatgatacgtcttgtagagcccaagccagacacgtacgtatgagatcccccact	8499
QY	766	TGCACACAGTGGCTCTCAGAAAGGGAGCCCTTGCCACAGACAGTGAATCCTTGGATGCTACCT	8255
Db	850	ttgggaagaggtcttactcttggtgcatcttc-tgaa-alcagc-tggagatcccaatccag	9060
QY	826	ACGGACGGTCTCTTCACTAGTGGCTCTTCATCAGACACACAGAGTGGGGCCCAACCCAG	8855
Db	907	gggaagagattccsagccggttcccaagagbgcaggaagaccctgagctactacagatat	9666
QY	886	GGTGTGGCATCTCCAGGCCCTTTCACCAAGAGAGGAGATGCTGGCTACTATTAAGTCT	9455
Db	967	ggagcttctcaaaagagctgaatgatcatgacctctcaacgagaaggtcccttgcta	1020
QY	946	GCTCTGTG--AAGGGGGCCACCAACAGAAATCCAGATCAACAAGTGCCTCTACATCT	1000
Db	1027	ccaagggcaaccagtggggtgggtatgagcacaagggagtggtcaaaaagaagttggt	1080
QY	1003	TCCGGGAACACCACTGGTGGCTTTGATGATGTGGAGAGCTTAAACCAAGTCAAGCT	1060
Db	1087	tccgtgaagagaagaagctgtgcagagagccatggtgtggtgcaactgtgattgatat	1144
QY	1063	ATCTGAAACCAAGAGGACTGGCGGGGCCATGTGCTGGCAGCTGACTGACTTATGATGACTT	1120
RESULT 10			
ID	MMD757	standard; RNA; ROD; 1506 BP.	
AC	D87757;		
NI	g1545818		
DT	15-SEP-1996 (Rel. 49, Created)		
DE	15-SEP-1996 (Rel. 49, Last updated, Version 1)		
DE	House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L precursor, complete cds.		
KW	ECF-L precursor.		
OS	Mus musculus (mouse)		
OC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
OC	Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.		
RN	1-1506		
RP	1-1506		
RA	Owhashi M.;		
RT	;		
RL	Submitted (06-SEP-1996) to the EMBL/GenBank/DBJ databases.		
RL	Macro Owhashi, The University of Tokushima, Faculty of Integrated		
RL	Arts and Sciences; 1-1 Mihama-Johsanjima, Tokushima, Tokushima 770,		
RL	Japan (E-mail:owhashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,		
RL	Fax:0886-56-7298)		
RP	[2]		
RP	1-1506		
RA	Owhashi M., Arita H., Taoka Y., Hayai N.;		
RT	"Molecular cloning and characterization of T lymphocyte-derived		
RT	eosinophil chemotactic factor (ECF-L)";		
RL	Unpublished.		
EH	Key	Location/Qualifiers	
EH	source	1..1506	

```

RESULT 10
ID MMD757 standard; RNA; ROD: 1506 BP.
AC D87757;
NI g1545818
DT 15-SEP-1996 (Rel. 49, Created)
DT 15-SEP-1996 (Rel. 49, Last updated, Version 1)
DE House mouse; Musculus domesticus bone marrow cells mRNA for ECF-L
DE precursor, complete cds.
KW ECF-L precursor.
OS Mus musculus (mouse)
OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
RN [1]
RP 1-1506
RA Owashi M.;
RT ;
RL Submitted (06-SEP-1996) to the EMBL/GenBank/DBJ databases.
RL Makoto Owashi, The University of Tokushima, Faculty of Integrated
RL Arts and Sciences; 1-1 Minami-Johsanjima, Tokushima 770,
RL Japan (E-mail:Ohashi@ias.tokushima-u.ac.jp, Tel:0886-56-7261,
RL Fax:0886-56-7298)
RN [2]
RP 1-1506
RA Owashi M., Arita H., Taoka Y., Hayai N.;
RT "Molecular cloning and characterization of T lymphocyte-derived
RT eosinophil chemotactic factor (ECF-L)";
RL Unpublished.
FH Key Location/Qualifiers
FT source 1..1506

```

```

FT      /organism="Mus musculus"
FT      /cell_type="bone marrow cells"
FT      /tissue_type="bone marrow"
FT      1..1197
FT      /gene="MECF-L"
FT      /codon_start=1
FT      /product="ECF-L precursor"
FT      /db_xref="PID:g1545819"
FT      /translation="MAKLILVGLAILNLVQSSYQLMCIYTTSNAKDRPIEGSFKPG
N      IDPCLCHLIYAFAGMKNNEITYTHEQDLRDYEAINGLKDKNTELKTLIAIGMKRFGP
A      SFSAMVSTPQNROIPIQSVIRFLROYNFDGLMDNQIPSSRSRSPDKHLFSLVKEM
R      KAFEESEVKDIPRLLLTSTGAGIIDIVISGRKIPELSGSLDYIOWMTYDLHPKRGY
T      GENSPLYKSPYDICKSADLNVDISIISYMKDHGAASEKLIVGFPAYGHPFILSDPSKTG
I      GAPTISTGPGKRYTDESGLAAYEVCFTFNEGATEVMDAPQEVPAAYQGENWVGIDNV
R      SFLKAQWLKDNNGAVVWRMDMDFSGSFCHQRHFPLTSLTKGDLNTHASCKRGPY
SQ      Sequence 1506 BP: 422 A; 337 C; 335 G; 412 T; 0 other;

Query Match      7.38; Score 121; DB 20; Length 1506;
Best Local Similarity 57.18; Pred. No. 1.14e-88;
Matches 609; Conservative 0; Mismatches 452; Indels 6; Gaps 4;

Db      70 ctgagtgtctactataccagttggcgaagagacccaatagaaggagttccaacct 129
Qy      96 CTGGTCTGCTACTTACCAACTGGCCCACTACAGACAGGGGAGGCTCCTCTCCGCC 155
Db      130 ggtbaatattgacccctgctgtgtactaccatgctatgctcttctggaatgcagat 189
Qy      156 AAGGACTTGGACCCCACTTTGACCCACTCATCTACCCCTTCCTGGACGACCAAC 215
Db      190 aatgagatcaactacacacatgagcaacttgcgtgacatgaagcattgaatgctgt 249
Qy      216 CACACACTGACACACACTGAGTGAATGAGACACTCTTACACAGGAGTCAATGGCCTG 275
Db      250 aaagacaagacacatgagcaaaacatctcctgagcattgagatggaattgagc 309
Qy      276 AAGAAATGAATCCCAAGCTGGAAGACCTCTTGACCTGGAAGCTGGAATTTCAACACT 335
Db      310 gccctgtcaagtcagtgctctactcctcagaacgcctcagatattcattcagta 359
Qy      336 CAGAAATTCAAGATATGAGACAGGCCACACACGCTGACACTTTGTCACTGGGCC 395
Db      370 atcaagttccttcgtcaataactttgagcctcaacctgagcagtcagtaacctg 429
Qy      396 ATCAGGTTTCTGGCAAAATACAGCTTTGAGGGCTTGACCTTGACCTGAGGAGTCCACAGA 455
Db      430 tctcgggaagagccctcctaagacaacatctcttcagtgctgtgtgaaggaatgc 489
Qy      456 AGCCAGGGAGGCCCTCCGCTAGACAGAGCGCTTCAACAACCTGTGACAGGACTTGGCC 515
Db      490 aagaacttttgaaggaatctgtggaagacattccaagcgtgtactactaccaca 549
Qy      516 AATGCTTCCAGCAGGAGCCAGACCTCAGGGAAGAGCGCTTCTGTGAGTGCAGCG 575
Db      550 gtagcagggaatctatgacgtaacatcaagctcgtgtacaaagctccctgaactgc 609
Qy      576 GTTCCAGCTGGGAGACCTATGTGATGCTGATAGCAGGTGACAAAACGCCCGCAAC 635
Db      610 ctggaataatactcagtcagacataatgctcattgacatcaggaatgagtcacag 669
Qy      636 CTGATTTTGTCAACTATAGGCTTACGACTTTCATGCGCTTCTGGAGAGAGGTCACAGGA 695
Db      670 gaaataagtcctcctataaactcattatgacattggaagagtcgtgactcaatgt 729
Qy      696 CATAAAGCCCTCTCAAGAGCAAGAGAGAGAGTGTGCACAGCAGCAGCTCAACGCTG 755
Db      730 gatcaatacttccatcaggaagacacatgagcagcttctgaagaagcattgtgg 789

```

```

Qy      756 GATGCTGCTGTGCAACAGTGGCTGCAAGAGGGAGACCCCTCCAGCAAGTGAATCTTGGC 815
Db      790 ttccacgatalgagacataccttaccctgagtgacccttcaagactgaatgtgtgc 849
Qy      816 ATGCTTACTAGAGAGCGCTTCCACAGTGCGCTCTCATACAGACACAGAGTGGGGGCC 875
Db      850 cctcaataagtaactggtcccaacaggaatgacacagatgatacaggaactcgtgtac 909
Qy      876 CCAGCCACAGGGGCTGTGCACTCCAGCCCTTACACCAAGAGAGAGGATGCTGGCCTAC 935
Db      910 tatgagttgtatcattcttgatgaagagacagcagtgagctcgtgagtcaccccgaa 969
Qy      936 TATGAAGTGTG--CTCTGGAAG--GG--GCCACCAACAGAGATCCAGATCTCAAG 989
Db      970 gtacccatgacctatcaggtgaatgagtggtgtgtatgacatgtcaagacattcaag 1029
Qy      990 GTGCCCTTACATCTTCGGGACCAACAGTGGTGGCTTTGATGATGTGAGAGACTTCMAA 1049
Db      1030 ttgaagctcagtggtcgaagacaacattagaagtgccgtgtgtcgtgcctgagac 1089
Qy      1050 ACCAAGTCACTATCTGCAAGCAGAGGACTGGCGGGGCCATGCTGTGGCAGCTGAC 1109
Db      1090 atgagtgaactcagtggtctcttctgtccacagagacattccctct 1136
Qy      1110 TTGATGACTTGTCCGCGCTTCTCTCAACACAGGCCGATACCCCT 1156

RESULT 11
LOCUS      D87757      1506 bp      mRNA
DEFINITION House mouse; Musculus domesticus bone marrow cells for ECF-L precursor, complete cds.
ACCESSION      D87757
NID      91545818
KEYWORDS      ECF-L precursor.
SOURCE      Mus musculus bone marrow cells cDNA to mRNA.
ORGANISM      Mus musculus
Eukaryote; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE      1 (bases 1 to 1506)
AUTHORS      Ohashi,M.
DIRECT SUBMISSION
TITLE      Submitted (06-SEP-1996) to the DDBJ/EMBL/Genbank databases. Makoto Ohashi, The University of Tokushima, Faculty of Integrated Arts and Sciences; 1-1 Minami-Johsanjima, Tokushima, Tokushima 770, Japan (E-mail:ohashieia.tokushima-u.ac.jp, Tel:0886-56-7261, Fax:0886-56-7298)
2 (bases 1 to 1506)
AUTHORS      Ohashi,M., Arita,H., Taoka,Y. and Hayai,N.
TITLE      Molecular cloning and characterization of T lymphocyte-derived eosinophil chemotactic factor (ECF-L)
JOURNAL      Unpublished (1996)
FEATURES
source      Location/Qualifiers
1..1506
/organism="Mus musculus"
/cell_type="bone marrow cells"
/tissue_type="bone marrow"
1..1197
/gene="MECF-L"
/codon_start=1
/product="ECF-L precursor"
/db_xref="PID:g1545819"
/translation="MAKLILVGLAILNLVQSSYQLMCIYTTSNAKDRPIEGSFKPG
NIDPCLCHLIYAFAGMKNNEITYTHEQDLRDYEAINGLKDKNTELKTLIAIGMKRFG
PASFSAMVSTPQNROIPIQSVIRFLROYNFDGLMDNQIPSSRSRSPDKHLFSLVKEM
KAFEESEVKDIPRLLLTSTGAGIIDIVISGRKIPELSGSLDYIOWMTYDLHPKRGY
GENSPLYKSPYDICKSADLNVDISIISYMKDHGAASEKLIVGFPAYGHPFILSDPSKTG
GAPTISTGPGKRYTDESGLAAYEVCFTFNEGATEVMDAPQEVPAAYQGENWVGIDNV
SFLKAQWLKDNNGAVVWRMDMDFSGSFCHQRHFPLTSLTKGDLNTHASCKRGPY"
BASE COUNT      422 a      337 c      335 g      412 t
ORIGIN

```





accession partial cds.  
 nid 017988  
 keywords 9602293  
 source  
 organism  
 Ovis aries  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.  
 1 (bases 1 to 1596)  
 reference Marshall,J.T., Brownlee,A.G. and Nancarrow,C.D.  
 authors Marshall,J.T., Brownlee,A.G. and Nancarrow,C.D.  
 title Cloning and sequencing of a cDNA encoding ovine oestrous glycoprotein  
 journal Reprod. Fert. Dev. 8 (2), 305-310 (1996)  
 reference 2 (bases 1 to 1596)  
 authors Marshall,J.T.A.  
 title Direct Submission  
 journal Submitted (02-DEC-1994) James T. A. Marshall, Division of Animal Production, Commonwealth Scientific and Industrial Research Organisation, Locked Bag 1, Delivery Centre, Blacktown, New South Wales, 2148, Australia  
 location/Qualifiers  
 FEATURES  
 source 1..1596  
 /organism="Ovis aries"  
 /strain="merino"  
 /sex="female"  
 /cell\_type="epithelial cell"  
 /tissue\_type="ampulla oviduct"  
 /dev\_stage="oestrus"  
 <1..1596  
 /codon\_start=1  
 /product="92kd oestrus-associated oviduct glycoprotein"  
 /db\_xref="pid:9602294"  
 /translation="LLVLKHKHDAHAKLYCFETMAFRSPGASILPRDPLCTHLYPFAASNNQIVKPKDLEKILPEPKKRNKRGKATLLSVGNMGFSRTKMLSTFNSRREFFNSVIALRTHEDPDGIDLEFLPGRSGARBMFFVLEELALAFKMAEQILMRRLILSAVSDPRVYOKATYARILGRLDISTLSIDHSMKRYGHNSPLFSPGDKSATSAMSTWROUGVPPERLNLGPLYGTFFHLRLASQELGAVAGPPEISGRTKQAGFLAYEVCSFVQRAKKRMINQVYPARKEWGYDAISFGKAFEEIKREFEGAMWTMLDLDFGRLNFCGTGFPVHTLNLNLVNDERSTSPSPFWSSTAVINSSRRPMPMTRLDTLGTGLILPGGAVAVTEHRKSATWTPRGATATPRPLSGRRRAPEKTESGPEKPLSVGHLANSPGIVAGVPHLDIGQKVPFGKACVPEVVTTSGRKATVPPDGAETLERL"  
 <1..>1596  
 /product="92kd oestrus-associated oviduct glycoprotein"  
 <1..36  
 /note="inferred from bovine sequence: GenBank Accession Number D16639"  
 /codon\_start=1  
 1..19  
 /note="derived from PCR primer based on bovine cDNA sequence: GenBank Accession Number D16639"  
 complement(1575..1596)  
 /note="derived from PCR primer based on bovine cDNA sequence: GenBank Accession Number D16639"  
 BASE COUNT 352 a 440 c 430 g 3/4 t  
 ORIGIN  
 Query Match 5.5%; Score 91; DB 47; Length 1596;  
 Matches 363; Similarity 58.5%; Pred. No. 9,79e+59;  
 Conservative 0; Mismatches 254; Indels 3; Gaps 2;  
 Db 40 aagcgggtgtgtattattacaacatgggattcagcggccggcgtcctcgcgacctg 99  
 Oy 93 AACACGTGTGCTACTTACCAACATCGGCCCACTACAGACAGGGGAGGCTGCTTC 152  
 Db 100 ccccgagatcgcagcccttctctcgcagccacccactggtatttgccttcctcgatgaac 159  
 Oy 153 CCCAAGGACTTGAGCCCGACGCTTTCACACCCACATCATCTAGCCCTTCGCTGCATGACC 212  
 Db 160 aacatcagatgtgttctctaagatgcccccgcggtatgagaataactctaccacagattcaac 219  
 Oy 213 AACCAACCACTTACAGACCACTGATGTAAGACGAGAC--TC-TCTACCAAGAGATTCAAT 269

D	b	220	aagctcaaggaggaagaacaggggctgtgaataaacctgtcttcgcttggtggggtggaattc	279
O	y	270	GGCCTGAAGAAGAAATCCCAACCTTAAAGACCCTGTATTACCATCGGAGGCTGGAAATTTC	329
D	b	280	ggcaacctgaagtccaccagaatgcgtgccacgttttccaaccgggaagaagtgtgcat	339
O	y	330	AGCACTCGAAGTTCCACAGATATGTGTGCCACGGCCACAACAGCGTGACACTTTGTCAAC	389
D	b	340	tcaagtatcgccctctgaaggaaacaatgcttgtgatgtgtcttgacctctctcttctac	399
O	y	390	TGCGGCATCAGAGTTCTTGCAGCAATATACAGCTTTGACGCGCCTTCACCTTGACTGGAGTAC	449
D	b	400	ccttgaaactcagaagccagccccgcgcgtgacgcgtctgacaccttctcttaacttgaag	459
O	y	450	CCAGGAACCCAGGGGAGCCCTGCGGTGACAGAGAGCCGTTCCAAACCTTGTTACAGAC	509
D	b	460	ctctcctgagagccttcaagaattagagcccagctcaccaatgctgcccagaagctgtcgtgt	519
O	y	510	TTGGCCATGCGCTTCCAGAGGAAGCCCAAGACCTCAGGGGAAGGAAGCGCTTCTTGAGT	569
D	b	520	gctgcgcgtctcttgggagacccccacgttatctcgaagaacatatgatagtcaogccttgygc	579
O	y	570	GCAGGGGTTCCAGCTGGGGCAGACTATGTGATGCTGGATACAGGTCAGCAAAAACGCG	629
D	b	580	agacacctgagcttcatcaagcgtctgtcttacgacttaacaggaagctggaagaatgc	639
O	y	630	CAGAACCTGGATTTTGTGCACACTTATGTGCTTAGCACTTCATGCTCTTGGGAGAGGTC	689
D	b	640	acaggacacaataagcctct	659
O	y	690	ACGGGACATAACAGCCCCCT	709
RESULT	LOCUS	15		
DEFINITION	BABEDSG	2228 bp	mRNA	PRI
ACCESSION	Papio hamadryas anubis estradiol-dependent oviduct-specific glycoprotein mRNA, partial cds.			11-OCT-1996
NID	M59903			
KEYWORDS	estradiol-dependent oviduct-specific glycoprotein.			
SOURCE	Papio hamadryas anubis cDNA to mRNA.			
ORGANISM	Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopithecinae; Papio.			
REFERENCE	Donnelly,K.M., Fazleabas,A.T., Verhage,H.G., Mavrogianis,P.A. and Jaffe,R.C.			
TITLE	Cloning of a recombinant complementary DNA to a baboon (Papio anubis) estradiol-dependent oviduct-specific glycoprotein Mol. Endocrinol. 5 (3), 356-364 (1991)			
JOURNAL	Medline			
REFERENCE	AUTHORS			
TITLE	2 (bases 1 to 2228)			
JOURNAL	Jaffe,R.C., Arias,E.B., O'Day-Bowman,M.B., Donnelly,K.M., Mavrogianis,P.A. and Verhage,H.G.			
REFERENCE	Regional Distribution and Hormonal Control of Estrogen-Dependent Oviduct Specific Glycoprotein mRNA in the Baboon (Papio anubis) Biol. Reprod. (1996) In press			
AUTHORS	3 (bases 1 to 2228)			
JOURNAL	Jaffe,R.C.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (05-MAR-1991) Randal C. Jaffe, Department of Physiology and Biophysics, University of Illinois College of Medicine, Chicago, IL 60612-7342, USA			
JOURNAL	location/Qualifiers			
FEATURES	source			
CDS	1..2228			
	/organism="Papio hamadryas anubis"			
	13..1884			
	/codon_start=1			
	/product="estradiol-dependent oviduct-specific glycoprotein"			
	/db_xref="PID:g1256951"			

/translation-"MKLLWGLVVLKHHDAHKLVCYFTNMAHSRPGPASTLPH  
LDLPEICHLIFAFASNNNOIVAKDLODEKLYPEPNKIKERNRELKTLISGIMNF  
GTSRFTTMLSTFANREKFIASVLSLRTHDEGDLDFLYPELRGSPMHDRTFLLI  
EELLEAFKRKELLIMRPRLLLSAAVGVPHIVQTSYDVAFGLRDLFINVLSYDLGS  
WEKFTGHSPLFSLPEDPKSSAYANMYRKLGAPESEKLIMGIPTGRFRLKASKNG  
LOATFICGASPGKTYKQAGFLAYFICSEVWGAKHMIWYOYVPYANKKEWVGDDA  
ISFSYKAMFIREHREHGMWTLDDVDVGRFPGCGRPFLYVMDIILVRAEFSSTL  
POFLSSAVNSSNDPRLAATKATTDIKILIPGCEAGVTEIHCKENMTTPRYTI  
VPTREIVSLGKHTVALGEKTEITGATMTISVGHOSMTPEKALTPVGHOSLPEKKT  
LTPVGHOSVTTGOKTLISVGHSHVTPGKTLTPVGHPSVTPSHQSVSPGQMTMPVH  
FOLETLRONTMAPRKAAVHEKVTVPSEKRIISVTPGQTVPLRGEYLTSETGTHPDG"

BASE COUNT 559 a 572 c 531 g 566 t  
ORIGIN

Query Match 5.5%; Score 91; DB 67; Length 2228;

Best Local Similarity 58.5%; Pred. No. 9.79e-59;

Matches 363; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

Db	79	aaactggtgtgtattatcccaactgaggacacagtcgagcccgcccgatcttg	138
Qy	93	aaactggtgtgtacttccacaaactggccacgtacacagggagggcttcctc	152
Db	139	cccctagaccctgacccctctctcaccacccatatttgcccttgcccaatgaac	198
Qy	153	cccagagacttgagcccccagccctttgcacaccatctatagcccttcgctg	212
Db	199	aacaatcagattgttgtaaggatctccagacgagaaaattctctaccagagtlcaac	258
Qy	213	aaccacacactgacacacactgagtggaatgacgagac--tctctacacagagttcaat	269
Db	259	aaactaaagagagagacagagagctgaaacacactgtctcactgagcggtggaactt	318
Qy	270	ggcctgaaagaaatgaaatcccaagaccctgttagccatcgagagctggaatttc	329
Db	319	ggcaccctcagggtccaccactatgtgtccacattgccaacggtgaaagttaattgtct	378
Qy	330	agcactcagaaagtctcacaatattgtagccacggcccaaacacgcttagacctttgtc	389
Db	379	tcaagtataatccctctcgaagacacatgaacttgatgcttgaatctttttctatat	438
Qy	390	tggccatcagagttctctgcgcaaatatagactttgacggccttgacctgac	449
Db	439	cttgactaagagcgagcccatgcatgacgggtggaacttctctcttaattgaag	498
Qy	450	ccaggaagccagggagcccttcgctgagacagagagccttcacaaacctgtgac	509
Db	499	ctccctgttgccttcggaagagagcgtctcaaccatgcycccaaggtgctgctct	558
Qy	510	ttggccaatgctcttcacagagagagccacgacctcagsgaagaaacgcttcttg	569
Db	559	gtctgctgttctcgtgaggtcccaacacatgcgtcaacatcctatgattgctg	618
Qy	570	gcagcggttcacagctgggacagacatgtagatgtagagtgagacaaatcgcc	629
Db	619	agactcctgatttcacatgctgtcttatgacttcacatggaagtgggaaagtcc	678
Qy	630	cagaaacttgattttgtcaaccttagagccttagacttccatggccttggaag	689
Db	679	acagacataatagcccccct	698
Qy	690	acgggacatracagcccccct	709

Search completed: Sun Jun 29 20:45:12 1997  
Job time : 1558 secs.







QY	109	TCNCCAACTGGGGCCCACTACAGAAAGGGGAGGCTGCCTCTCTGCCCCAAGACTTGGACC	168
Db	214	gcttcctctgtaccacaatcatctacacgtcttgcacaatataaagcaacgatcatcatgaca	273
QY	169	CCAGGCTTTGGACCCACCACCTCATCTACCGCTTCGGTGGACGACCAACACACACTGAGA	228
Db	274	cctggagatgaaatgatatgtacgtctcaaggcatgctccaacaacatccaacaagaaacc	333
QY	229	CCACTGATGGAAATGACAGACTCTTCACACAGAGATTCAATGGCCCTGAAGAAATATAATC	288
Db	334	ccaactccaagaactctctgtctgtcgagagatgaaacttggcttgcacaagaatttcca	393
QY	289	CCAACCTAACAACCTGTTACCACTCGGAGGCTGGAAATTTCACACTCAGAAATTACAG	348
Db	394	agatagcctccaacaaccagaagtcgcgcgaacttcatcaagaigtacagcccatcttgc	453
QY	349	ATATGTTAGCCACGCGCCACAAACCGTGCAGACCTTTGTCACATCGGCCATCAGGTTCTGC	408
Db	454	gcaccacatggtttgatggggtgtaccttgcctcgtctacacttga----ggaag---	505
QY	409	GCAAAATACAGCTTTGACGGCCTTTACCTTGACTGGGAGTCCACGAGAAGCCAGGAGAGCC	468
Db	506	-----agacaacacacatttccaccaccatcaacaagaatgaaagccgaattatata	558
QY	469	CTGCGGTGAGACAAGAGGCGCTTCACAACTCGTACAGAGACTTGGCCAAATGCCCTTCCAC	528
Db	559	aggaagcccaag--c--cagggaaaaagcagctcctcgtctcaagcgaacatgctcgcggga	615
QY	529	AGGAAGCCCAACCTCAGGGAAGGAAAGCGCTTCTTGATGTGCAGGCGGTTCACGTGGGC	588
Db	616	aggtacacattgacagcagcgtatgaacttgcacaagatatcccaacaccttggatttcat	675
QY	589	AGACCTATGTTGAGTCTGTGATACAGAGGTGACAAATAAGCCGCCAAGCCTGGATTTGTCA	648
Db	676	gcatcatgacctacatatttcatgtgcgcgtgcgtggagaccaagccatcacagtcgcc	735
QY	649	AACCTATGCGCTACGACTTCATGTGCTTGTGGAGAAGGTACAGGGACATTAACAGCCCC	708
Db	736	tgttcggaagtcagagagatgcaagtcctgaacagatccaacacactgacatagctgtgtg	795
QY	709	TCTAACAAAGGCAAGAGAGAGTGGTGACAGACGACACCTCAACGHTGAGTGTGTGTC	768
Db	796	ggtacatgtttagcgtcgggggtctccctgcacgtaaagctggtgtagtgggcatcccaacttg	855
QY	769	AACACTGGCTCCAAAGGGGAGCCCTCCACAGCAAGCTGATCCCTTGCGATGCCCTACCTACG	828
Db	856	ggaagagcttaactctgtgctctctc--tgaagctg--gtgttcagagcccaatcagagac	912
QY	829	GACGCTCTTTCACACTGGCCCTCTCTCATACAGACACACAGATGGGGGCCCCAGCCACAGGAT	888
Db	913	cgggaattccagcgcggtttcaccaaagagcgaaggacccttgcctactatagatctgtg	972
QY	889	CTGGCACTCCAGGCCCTTTCACCAAGGAAGGAGATGTGGCTACTATGAAGTGTGCT	948
Db	973	acttctccgcggagcccaagtcataagaacctcggcgaagcaggttccctatgccaaca	1033
QY	949	CCTGGAAAG--G--G--GCCACCAAAAGAGAAATCCAGGATTCAGAAAGTGGCCCTACATTTCC	1001
Db	1033	aggggaacacagttggttagatagacagacaacgaagacgftcaaaagcaaggtgcagtacc	1097
QY	1006	GGGACACCACTAGTGGGTGGCTTTATATATGTGGAGACTTCAAAACCAAGTAACTATAC	1066
Db	1093	tgaagatagacgactgtgcagcgcgcatagtatagagccctcggaactgtagtacttcag	1155
QY	1066	TGAAGCANAAAGGACTGGGCGGGGCCATGATGTGGGACACTGGACTTAAGATGATTTGGCG	1122
Db	1153	gctcctctgcgcgcacag 1170	
QY	1126	GCTTCTCCTGCACACAG 1143	

AC	N81756; 19-Oct-1990 (first entry)
DE	Gene encoding polypeptide involved in protective mechanisms
KW	Immune response; cell growth; ss.
FH	Key Location/Qualifiers
FT	CDS 1..966
FT	/*tag= a
FN	J63032898-A.
PD	30-JAN-1988.
PE	16-JUL-1986; 167518.
PR	16-JUL-1986; JP-167518.
PA	(DAIN) Daiinippon Pharm KK.
DR	WPt: 88-068419/10.
P	P-PsDB: P61342.
PT	New polypeptide and DNA encoding it - related to protective mechanisms such as immune response etc.
PS	Disclosure; P: Japanese.
CC	Polypeptide encoded is involved in protective mechanisms such as immune sequence, cell growth and activation of protective functions.
SQ	Sequence 966 BP; 238 A; 282 C; 247 G; 199 T;
Query Match 7.9%; Score 131; DB 1; Length 966; Best Local Similarity 66.1%; Pred. No. 2,21e-69;	
Matches	269; Conservative 0; Mismatches 138; Indels 0; Gaps 0.
D <sub>b</sub>	23 caggcttggtgctcctggctgcctccagtygtcgtctgcatacaaaagtgtgctact 82 
O <sub>y</sub>	49 CAGGTTTCATGGTCTCGTGATGATCCCATGTGGGCTCTGTCGAAMAAGTGTGTA 108 
D <sub>b</sub>	83 aaacagcttgtcccagtaaccgggaagcgatggagctgtcttcaccaatgacctaac 142 
O <sub>y</sub>	109 TCACCACACTGGGCCCGATGACAGAGAGGGGGAGGCTCCTCTCTGCCAAGACCTTGGAC 168 
D <sub>b</sub>	143 gttcctctgtcacccacaatactacagcttgccaatatagaacgatcatcatcgaca 202 
O <sub>y</sub>	169 CCAGGCTTTGCACCCCACTCATCTACAGGCTTTCGCTGGCATGACCACCACCTGTGACA 228 
D <sub>b</sub>	203 cctgggaattggaatgatgtgaagctctacagctcaggtcctcaacacactcaagaacgagAAC 262 
O <sub>y</sub>	229 CCACGAGTAGGAATGACGAGACTCTCTCACGAGAGTTCAATGGCCTGAGAAGATGATC 288 
D <sub>b</sub>	263 ccaacttaaaagactctctgtctgtcgtcggagatgtgaacttgggtctccaagaattttcca 322 
O <sub>y</sub>	289 CCMACTGAAAGACCTGTTAGCCATCGGAGGCTGGAATTTCAGCACTCAGAAAGTTTCACAG 348 
D <sub>b</sub>	323 agatagctccaacaaccagagagtcgcggacttcatacaagtaagiagtcgccattcttcgc 382 
O <sub>y</sub>	349 ATATGTTAGCCACGCGCCACAACCGTCAGACCTTTGTCAACTCGGCCATCAGGTTTCTGC 408 
D <sub>b</sub>	363 gccaccatgcttgattgatgggttgagacttcgctgagctctaactctgca 429 
O <sub>y</sub>	409 GCAAATACAGCTTTGACGCGCTTGACCTTGACTGGAGGATACCACGAGA 455 
RESULT 3	
ID	Q10572 standard; DNA; 1047 BP.
AC	Q10572;
DT	09-APR-1991 (first entry)
DE	Human Natriuretic Peptide Receptor B.
KW	NPBR; ANP; BNP; CNP; Kidney failure; heart failure; protein kinase;
KM	hyperaldosteronism; glaucoma; guanyl cyclase.
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	Peptide 1..22
FT	/label= signal sequence
FT	Protein 12
FT	/label= mature NPBR
FT	Domain 23..455
FT	/label= extracellular domain
FT	/note= "binds natriuretic peptides A,B and C]"
FT	Domain 456..456
FT	/label= transmembrane domain
FT	Domain 479..1047

```
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified -site 24..26
FT /label= N-glycos_-site
FT Modified -site 35..37
FT /label= N-glycos_-site
FT Modified -site 161..163
FT /label= N-glycos_-site
FT Modified -site 195..197
FT /label= N-glycos_-site
FT Modified -site 244..246
FT /label= N-glycos_-site
FT Modified -site 277..279
FT /label= N-glycos_-site
FT Modified -site 349..351
FT /label= N-glycos_-site
FT Modified -site 600..602
FT /label= N-glycos_-site
PN MO9100292-A.
PD 10-JAN-1991.
PF 22-JUN-1990; 003586.
PR 23-JUN-1989; US-370673.
PA (GENE ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB: Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B. NPRO, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRO analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRO can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 5.6%; Score 92; DB 2; Length 1047;
Best Local Similarity 10.6%; Pred. No. 1,24e-42;
Matches 106; Conservative 279; Mismatches 603; Indels 15; Gaps 14;

Db 5 sunnvnanaagvyrngarntlnavnnhnsyawaavrignavanavangrannvdh 64
QY 120 GCCCAGTACAGACGAGGGGAGGCTTCCTCCAGAGACTTGAGCCCACTGTTG 179
Db 65 rrvssnnngacsnynnaaeavdknyhndnnngngcynnaas-varnashtvnnmtag 123
QY 180 ACCGACCTCATCTACGCTTCGCTGCGATGACCAACACACAGCTGAGCAGCTGAGTG 239
Db 124 avasgnakndhyrtvrtgnsankngnnvtnhgnnwraaraanyndaridrhnynt 183
QY 240 AATGACAGAG-ACCTC-C-TACCGAGATTCAATGCCCTGAAGAAGTGAATCCCAAGCTG 296
Db 184 ngvnnannngsnsvnhuyarnngnnnathnnrangrtnvnycgnnmmhnnnnan-r 242
QY 297 AAGACCTGTATACGATCGAGGCTGGAATTTACGACACTCAGAAATTCACAGATATGTA 356
Db 243 nmtngdyrnnnyndvngnsrtaagtatgtnvndttrnnanannantvnnvtrynn 302
QY 357 GCCACGCGCAACACCTGACAGACCTTGTCAACTCGGCCCATCGGTTTCTGCCCAATAC 416
Db 303 nnnynnnnnrnnrrndngvngsnmmnngcnydgmmnyannvnnntnngttrndp 362
QY 417 AGCTTTGACGGCCTTGACCTTGCTGGAGATACCCAGAGAACCCAGGGAGCCC-TGCCGT 475
Db 363 rrvnknngrrtyhvtgnvnmvdkndirntdvrvwamgdndsddnnaahysganknmwtg 422
QY 476 AGACAGAGAGGC-GCTTCACAAACCTGTGATACAGGACTTGCCCATGCTTCACAGAGAG 534
Db 423 rnnn-vv-kgannsdnnncandndnscktlmstnanvngntnmmngvsnmmnrk 480
```

```
QY 535 CCCAGACCTCAGGAGAGAACGCGCTTCTCTGAGTCCAGCGGTCCAGCTGGGACACCT 594
Db 481 mnnknnaasmvnrwrwnnnngnsnyhkgagsrtnsnrgssysnmfahkyannant 540
QY 595 ATGTGATGC-TGGATACAGAGTGAGGACAAATCGCCCAAGCTGATTTTGCAACCT 653
Db 541 ghkngrvannknyknkrnntrvnnnnkhmrdivnnnhltrngacndmnnenvtnyc 600
QY 654 -ATGGCTACGACTTCCATGGCTTGGGAGAGAGTACAGGACATPAAACAGCCCTCTA 712
Db 601 rgsnnndnnndsnndmwmrysnndnvkymannhnsnshgsknsrvcvdsrvnkh 660
QY 713 CAAGAGCCAGAAAGAGAGTGTGACAGCACAGCCCAACAGTGAGTGTGTGACACA 772
Db 661 tdynasnrsstannddnnaanyaknntannsnngnnnttgmnaadvysngnnnnanr 720
QY 773 GTGGCTGCAGAAAGGGGAGCCCTGCCAGCAAGCTGATCCTTGCGATGCCCTACAGGAG 832
Db 721 sgnyngnndnanknnvknrgnrnynrnsndrtlnnnnnvnnmtrcwandnandrng 780
QY 833 CTCCTTCACTGCGCTCCTCATCAGACACAGAGTGGGGGCCCCAGCACAGGCTGTG 892
Db 781 mkgmrrrnkngtlesndnnnnnnnnnnnnnnkvvnrtnaaynnkrranaanynnmh 840
QY 893 CACTCCAGGCCCTTCCACAGAGAGAGAGTGTGCGCTGATATAGTGTGTCTGTG 952
Db 841 svannknrgntvananadvtnysndvgtlansanstnmvntnndtycnd--and 898
QY 953 GAAAGGGGCCACCAACACAGAAATCCAGATCAGAAAGTGCCCTTACTTCCGGACAA 1012
Db 899 ndvkvntngdnyamvsgnrgnrnhanmananadvssnrnrhndnnrn 958
QY 1013 CCAAGTGGTGGCTTGATGATGTGAGAGCTTCAAAACCAAGTGCATCTGTAAGCA 1072
Db 959 gvtgtvncagvynknmrycnngdvtntsrmsngnankhv 1001
QY 1073 GAAGGAGCTGGCGG-GGCCATGTGC-TGGCAGCTGAGCTTAG 1113

RESULT 4
ID Q90443 standard; cDNA to mRNA; 2504 BP.
AC Q90443;
DT 10-JAN-1996 (first entry)
DE Murine oviduct specific glycoprotein cDNA.
KW Murine oviduct specific glycoprotein; recombinant production;
KW MOP: ds.
OS Mus musculus.
FH key Location/Qualifiers
FT CDS 3..2160
FT /tag= a
FT /note= "START codon absent"
FT sig_peptide 3..56
FT /tag= b
FT mat_peptide 57..2156
FT /tag= c
PN J07107979-A.
PD 25-APR-1995.
PF 15-AUG-1994; 214227.
PR 19-AUG-1993; JP-227881.
PA (KINO-) KINOSSET PEPTIDE KENKYUSHO KK.
DR WPI: 95-190179/25.
DR P-PSDB: R90443.
PT New DNA encoding an oviduct-specific glycoprotein - useful for
PT recombinant protein production in high quantities.
PS Claim 1: Pages 11-14; 22pp; Japanese.
CC Q90443 encodes R73997 murine oviduct specific glycoprotein (MOP).
CC The cDNA can be used for the commercial recombinant prodn. of
CC MOP in high quantities.
SQ Sequence 2504 BP; 678 A; 616 C; 538 G; 672 T;

Query Match 5.5%; Score 91; DB 15; Length 2504;
Best Local Similarity 58.5%; Pred. No. 5.83e-42;
```

Matches 363; Conservative 0; Mismatches 254; Indels 3; Gaps 2;

Db 60 aaactggtgtgctattccattccattgagacatactgagccgacctgctccatcatg 119  
 OY 93 AAACGTGCTGTGCTACTTCCACCACTGGGCCAGTACAGCAGGGGAGGCTGCTCCG 152

Db 120 ccccatgacgtgagccctctcttctgtaacatctgataattgcttgcctcaatgagc 179  
 OY 153 CCCAAGGACTTGGACCCAGCTTGGACCCACCTCATCTAGCCCTGCTGGCATGACC 212

Db 180 aacatcagattgtgtccaaagatctgcagatgaaacgttctctatccagatcaac 239  
 OY 213 AACCAACGACT-GAG--CACCACTAGTGAATGAGAGAGACTCTTACCGAGGATTCAT 269

Db 240 aaactcaagaaagaaacagagagagctgaaacacgtctgcatctgagaggtggaacttc 299  
 OY 270 GGCCCTGAAGAAATGAATGCCAAGCTGAAGACCTGTAGCCATGGAGAGCTGGAAATTC 329

Db 300 ggcacatacaggttaccgctatgtctccacacttgccaacgagaaagtattatgac 359  
 OY 330 AGCACTCAGAAAGTTACAGATATGATGTAGCCACGCGCAACACCGTACAGACTTGTCAAC 389

Db 360 tcaagcatatcctctctgtaaatatagattgtgtctgaccttctctctgtac 419  
 OY 390 TCGGCAATGAGTTCTTGCGCAAAATACAGCTTTGACGGCTTGACCTTGACTGGAGATAC 449

Db 420 cctgacctacagagagcagctcccccacagaccggtgaaattctctcttcttaataagag 479  
 OY 450 CCAGAAAGCCAGGAGAGGAGCCCTCCGTAGCAAGAGCCGTTACAAACCTGTGTACAGAC 509

Db 480 ctccagattgctctcagagagagagccctgtcactcagcaccacagaggtgctgtac 539  
 OY 510 TTGGCCAATGCTCTTCAGAGAGAAACCCAGACCTCAGGGAAGAAAGCGCTTCTGTAGT 569

Db 540 gctgtgtctctgagatcccaagcatatccatcactgcttcttaagagcgtctcttaaga 599  
 OY 570 GCAGGGTTCACGCTGGCGACAGCTATGTGATGCTGATACAGAGTGAGACAAATCGCC 629

Db 600 aggcgttggattcatatgattgtctcttaatacttaccatgagagttggaagatt 659  
 OY 630 CAGAACCTGGATTGTGTCAACCTTATGGCTTACGACTTCAGTGCCTTTGGAGAAAGTTC 689

Db 660 acaagacacaaatagctctt 679  
 OY 690 ACGGACATTAACAGCCCTT 709

RESULT 5  
 ID 090442 standard; cDNA to mRNA; 1994 BP.  
 AC 090442;  
 DT 10-JAN-1996 (first entry)  
 DE Bovine oviduct specific glycoprotein cDNA.  
 KW Bovine oviduct specific glycoprotein; recombinant production;  
 KM BOP; ds.  
 OS Bos taurus.  
 FH Key Location/Qualifiers  
 FT CDS 2..1615  
 FT /\*tag- a  
 FT sig-peptide 2..55  
 FT /\*tag- b  
 FT mat\_peptide 56..1612  
 FT /\*tag- c  
 FT J07107979-A.  
 PN 25-APR-1995.  
 PD 15-AUG-1994; 214227  
 PR 19-AUG-1993; JP-227881.  
 RA (KINO-) KINOSEI PEPTIDE KENKYUSHO KK.  
 DR MPI; 95-190179/25.  
 DR P-PSDB; R73991.  
 PT New DNA encoding an oviduct-specific glycoprotein - useful for  
 PT recombinant protein production in high quantities.  
 PS Claim 1; Pages 7-9; 22pp; Japanese.

CC 090442 encodes R73991 bovine oviduct specific glycoprotein (BOP).  
 CC The cDNA can be used for the commercial recombinant prodn. of  
 CC BOP in high quantities.  
 SO Sequence 1994 BP; 445 A; 536 C; 520 G; 493 T;

Query Match 5.0%; Score 83; DB 15; Length 1994;  
 Best Local Similarity 57.9%; Pred. No. 1.33e-36;  
 Matches 359; Conservative 0; Mismatches 258; Indels 3; Gaps 2;

Db 59 aagctggtgttatttaccacaaatgtgacatcagtgagcccgacctgctcaatctg 118  
 OY 93 AAACGTGCTGTGCTACTTCCACCACTGGGCCAGTACAGACAGGGGAGGCTCGCTCTG 152

Db 119 cctcggaactggaacccgttctctgacccacctggtatgtgcttgcctgatgagc 178  
 OY 153 CCCAAGACTTGGAGACCCAGGCTTTGGACCCACCTCATCTAGCTTGCCTGGCATGACC 212

Db 179 aacatcagattgttcttaagatcccaagatgagaatactctaccacagattcaac 238  
 OY 213 AACCAACGACTGAGCAGCAGCTAGTGAATGACAGAGC--TC-TCTACAGAGATTCAAT 269

Db 239 aagctcaagagagaaagagagagctgaaacgctgtctcaccagggggtggaacttt 298  
 OY 270 GGCCCTGAAGAAATGAATCCCAAGCTGAAAGACCTGTTAGCCATCGAGGCTGGAATTC 329

Db 299 ggcacgtgaggttaccacagatgtctccacgttttccaacgagaaaggtctgacgt 358  
 OY 330 AGCACTCAGAAATGATACAGATATGATGTAGCCACGCGCAACACCTGTGTACAGAC 389

Db 359 tcaagtatgcctctctgagacacatggttgaatgtgtctgacaccttctctgtac 418  
 OY 390 TCGGCAATGAGTTCTTGCGCAAAATACAGCTTTGACGGCTTGACTGTGATGGAGATAC 449

Db 419 cctgacctcaagagagccctgctgctgacgctgagacgttctctcttcttaagaag 478  
 OY 450 CCAGAAAGCCAGGAGAGGAGCCCTCCGTAGCAAGAGAGCGCTTCACAAACCTGTGTACAGAC 509

Db 479 ctccgtcagaccttcaagaatgagccagctaccatgcgtccaaagctgtcgtgct 538  
 OY 510 TTGGCCAATGCTCTTCAGAGAGAAACCCAGACTTCAGGGAAGAAAGCGCTTCTGTAGT 569

Db 539 gctgcgtctctgagagaccccatgtctgccaagaagcgtatgaagcagcgtctctgagc 598  
 OY 570 GCAGGGTTCACGCTGGCGACAGCTATGTGATGCTGATACAGAGTGAGACAAATCGCC 629

Db 599 agactccgtgattcatatgattgtctcttaatacttaccatgagagctggaagatt 658  
 OY 630 CAGAACCTGGATTGTGTCAACCTTATGGCTTACGACTTCAGTGCCTTTGGAGAAAGTTC 689

Db 659 acaagacacaaatagccctct 678  
 OY 690 ACGGACATTAACAGCCCTT 709

RESULT 6  
 ID 010572 standard; DNA; 1047 BP.  
 AC 010572;  
 DT 09-APR-1991 (first entry)  
 DE Human Natriuretic Peptide Receptor B.  
 KW NPB; ANP; BNP; CNP; Kidney failure; heart failure; protein kinase;  
 KW hyperaldosteronism; glaucoma; guanyl cyclase.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /label- signal sequence  
 FT Protein 12  
 FT /label- mature NPBR  
 FT Domain 23..455  
 FT /label- extracellular domain  
 FT /note- "binds natriuretic peptides A, B and C"  
 FT Domain 436..456  
 FT /label- transmembrane domain  
 FT Domain 479..1047



OY 93 AACTGCTCTCTACTTACCAACTGGGCCAGATACAGAGGGGAGGCTCTCTG 152  
DB 120 ccccgtagcttgatccctctcttgtagacacagctgattgcttgccttcgagtgagc 179  
OY 133 CCCAAGGACTTGGACCCACGCTTTGGACCCACCTCATCTACGCTCTGCTGGCATGACC 212  
DB 180 aacaatcagattgttgcacataatctccagagtagaagaatactctatccagattcaac 239  
OY 213 AACCAACAGACTGAGCACCACCTAG-TGA-ATGACGAGACTCTCTACAGAGATTCAAT 269  
DB 240 aacatcaagagaagagaagagagccctgaaacactactctgttggagagctggaactc 299  
OY 270 GGGCTGAAGAAATGATCCCAAGCTTGTGAGACCTGTGACCATGAGAGCTGGAATTC 329  
DB 300 ggcacatcacggttcacacatctgctgcaccccttgccagcgcgtgaaattttatgt 359  
OY 330 AGCACTCAGAAATTCACAGATATGGTAGCCAGCCACACACCGTCAGCTTTGTCAAC 389  
DB 360 tcaattgtatctctctggaacacatgcttggatggcttgcctctctctgtac 419  
OY 390 TCGGCATAGGTTCTCTGGCAATATACAGCTTTGACGGCTTGACCTTGAGTGAGATAC 449  
DB 420 cctgtagctcagagcagccc 439  
OY 450 CCAGGAAGCCAGGGAGGCC 469

## RESULT 8

ID N81164 standard; DNA: 204 BP.  
AC N81164.  
DE Base substituted E.coli beta-galactosidase alpha-fragment.  
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.  
OS Escherichia coli.  
FH Key Location/Qualifiers  
FT misc\_feature 19..69  
FT /tag- a  
FT /function-multiple cloning site  
FT primer\_bind 187..204  
FT /\*tag- b  
PN EP-285123-A.  
PD 05-MAY-1988.  
PR 30-MAR-1988; 105163.  
PR 03-APR-1987; US-034819.  
PA (SUSO) SUDOMEN SOKERI OY.  
PI Lehtovaara P, Knowles J, Kolivula A, Bamford J, Reinikainen T.  
DR WPI: 88-279927/40.  
PT Introducing random point mutations into nucleic acids -  
PT by prepn of single stranded template, annealing a primer, elongation,  
PT misincorporation, completion of molecules and screening.  
PS Disclosure; P: English.  
CC Random point mutations were introduced into the alpha fragment of  
CC E.coli beta galactosidase. The wild type sequence was obtained as a  
CC single stranded template and an oligonucleotide was hybridised to  
CC it to generate a popn of DNA molecules which terminate at all  
CC possible nucleotide positions within a specified region. The  
CC variable 3' ends generated in this way are used as primers for  
CC reverse transcriptase. Nucleotides are misincorporated by the  
CC transcriptase and the molecules are completed to forms that can be  
CC amplified and then expressed in a suitable host-vector system.  
CC The sequence covers all 176 diff base substitutions, most of which  
CC occurred singularly in any given mutant.  
CC See also P80575.  
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

## Query Match

Best Local Similarity 3.1%; Score 51; DB 1; Length 204;  
Matches 18; Conservative 62; Mismatches 36; Indels 1; Gaps 1;

DB 75 yydcdhvycgymrtthhrrmbhvryrdynsdaaayccyrsvkxycdcynachhd 134  
OY 628 CCCAGAACCTGGATTGTGCAACCTTATGCGCTACAGACTTCATGCGCTTGGAGAAAG 687

DB 135 hyvbbvynvnhnncncobnhvchvbnhbnhrwavyrhdarriddhvcvchc 191  
OY 668 TCACGGGACATACAGACCCCTC-TACAAGAGGCAAGAGAGGTGTCAGCAGCC 743

## RESULT 9

ID Q51746 standard; cDNA: 91 BP.  
AC Q51746.  
DE Oligonucleotide probe MK14-A  
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
OS Synthetic.  
PN EP-571911-A.  
PD 01-DEC-1993.  
PR 24-MAY-1993; 108325.  
PR 26-MAY-1992; US-889651.  
PA (BECT ) BECTON DICKINSON CO.  
PI Shank DD, Spears PA;  
DR WPI: 93-378844/48.  
PT New oligo:nucleotide probes specific for Mycobacteria - used for  
PT detection and amplification of Mycobacteria nucleic acid in  
PT samples  
PS Claim 3; Page 14; 23pp; English.  
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
CC cross reacted to a few non-mycobacterial spp. The probe may  
CC be useful as an initial screen for mycobacterial infection.  
CC See also Q51735-45 and Q51747-59.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

## Query Match

Best Local Similarity 2.7%; Score 45; DB 9; Length 91;  
Matches 1; Conservative 47; Mismatches 3; Indels 0; Gaps 0;

DB 10 gsvhsyyvvvhhshvhhvhhvhhvsvvvvhhvhhvhhvhyvsv 60  
CP 1386 GGCCTGCGGAGCTTGTGCGAAGACGCGCCCGCTGACAGCTGAGA 1336

## RESULT 10

ID Q51746 standard; cDNA: 91 BP.  
AC Q51746.  
DE Oligonucleotide probe MK14-A  
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
OS Synthetic.  
PN EP-571911-A.  
PD 01-DEC-1993.  
PR 24-MAY-1993; 108325.  
PR 26-MAY-1992; US-889651.  
PA (BECT ) BECTON DICKINSON CO.  
PI Shank DD, Spears PA;  
DR WPI: 93-378844/48.  
PT New oligo:nucleotide probes specific for Mycobacteria - used for  
PT detection and amplification of Mycobacteria nucleic acid in  
PT samples  
PS Claim 3; Page 14; 23pp; English.  
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
CC cross reacted to a few non-mycobacterial spp. The probe may  
CC be useful as an initial screen for mycobacterial infection.  
CC See also Q51735-45 and Q51747-59.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

## Query Match

Best Local Similarity 2.5%; Score 42; DB 9; Length 91;  
Matches 2; Conservative 45; Mismatches 5; Indels 0; Gaps 0;

DB 12 svhsyyvvvhhshvhhvhhvhhvsvvvvhhvhhvhhvhyvsv 63  
OY 149 CCGCCCAAGACTTGACCCAGCGCTTGCACCCACCTCATATACAGCCTTC 200











\*\*\*\*\*  
MUSEE  
(TM)  
\*\*\*\*\*

Release 2.1d John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligentics, Inc.

Mparch\_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Jun 29 20:49:03 1997; Maspar time 1172.65 Seconds/sec

Tabular output not generated.

Title: >US-08-663-618A-3  
Description: (1-1656) from US08663618A.seq  
Perfect Score: 1656  
N.A. Sequence: 1 GCCTGACCTGCGCGCTGAC.....TGTGCCCCCTCAAAAAAAA 1656  
Comp: CGACGCTGCGACGGCGACTCG.....ACACGGGAGGATTTTTTTTT

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 886179 segs, 320143548 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

EST-STS

1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99

Database:

EST-STS-TWO

100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

Statistics: Mean 11.532; Variance 2.180; scale 5.290

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	345	20.8	428	151	YE31G05.s1 Homo sapie	0.00e+00
2	245	14.8	151	T94272	YE31G05.s1 Homo sapie	0.00e+00
3	215	13.0	241	HS6124	human STS SHGC-17364	0.00e+00
4	215	13.0	241	G29124	human STS SHGC-17364	0.00e+00
5	79	4.8	405	41	YP05A12.r1 Homo sapie	1.44e-100
6	78	4.7	541	113	Y29C01.r1 Homo sapie	1.34e-98
7	78	4.7	541	207	Y29C01.r1 Homo sapie	1.34e-98
8	73	4.4	533	144	YC24B03.s1 Homo sapie	7.80e-89
9	73	4.4	533	200	human STS WI-14165	7.80e-89
10	71	4.3	336	159	Y27B09.r1 Homo sapie	5.83e-85
11	71	4.3	336	159	Y27B09.r1 Homo sapie	5.83e-85
12	66	4.0	266	35	Y142F01.r1 Homo sapie	2.27e-75
13	65	3.9	396	134	Y179C12.r1 Homo sapie	1.81e-73
14	60	3.6	306	215	ma47G07.r1 Soares mou	4.65e-64
15	60	3.6	306	144	Y199G05.s1 Homo sapie	4.65e-64
16	60	3.6	421	139	Y199G05.s1 Homo sapie	4.65e-64
17	60	3.6	438	76	Y27B09.s1 Homo sapie	4.65e-64
18	59	3.6	586	170	md19C03.r1 Soares mou	4.65e-64
19	59	3.6	589	133	mg03C04.r1 Soares mou	3.38e-62
20	56	3.4	320	144	YC24B03.s1 Homo sapie	1.17e-56
21	49	3.0	299	151	YE25B05.s1 Homo sapie	4.86e-44
22	43	2.6	459	31	Y27B09.s1 Homo sapie	1.22e-33
23	41	2.5	461	56	Y25B05.s1 Homo sapie	2.80e-30
24	38	2.3	392	95	Y25B05.s1 Homo sapie	2.36e-25
25	36	2.2	335	47	Y231C01.s1 Homo sapie	3.71e-22
26	35	2.1	380	96	C.elegans CDNA clone	1.38e-20
27	35	2.1	380	96	C.elegans CDNA clone	1.38e-20
28	34	2.1	360	47	PK05A03.s1 Caenorhabd	1.38e-20
29	34	2.1	360	45	C.elegans CDNA clone	4.86e-19
30	34	2.1	429	44	C.elegans CDNA clone	4.86e-19
31	33	2.0	360	45	C.elegans CDNA clone	1.53e-17
32	33	2.0	360	47	C.elegans CDNA clone	1.53e-17
33	32	1.9	323	16	mm24C01.r1 Striatagene	5.19e-16
34	32	1.9	323	24	mm24C01.r1 Striatagene	5.19e-16
35	31	1.9	379	224	C.elegans CDNA clone	1.56e-14
36	31	1.9	379	224	C.elegans CDNA clone	1.56e-14
37	28	1.7	280	231	zm02H01.r1 Striatagene	2.89e-10
38	28	1.7	280	14	zm02H01.r1 Striatagene	2.89e-10
39	26	1.6	331	11	SMNFCAB00SK Brugia ma	1.40e-07
40	26	1.6	331	223	SMNFCAB00SK Brugia ma	1.40e-07
41	25	1.5	447	200	human STS WI-14048	2.72e-06
42	25	1.5	447	109	Y280F04.s1 Homo sapie	2.72e-06
43	24	1.4	375	185	md98B05.s1 Soares mou	4.82e-05



Db	299	ctcttttgcgnttcctcttancacaggtttttgtctcagcttgcctttt	358
Oy	1532	ctttttt-gtggcttt-cccttta-tccagcctttct-gctt-cagccttgcccttctttt	1586
Db	359	ttnttggtttcttcggggttcgccctttcattingcaantaat	403
Oy	1587	ttctt-gggcttccttgccgct-gcccttttcactt-gcaaaatnaat	1628
RESULT	3	standard; DNA; STS; 241 BP.	
ID	HSG124.		
AC	G29124.		
NI	g1592433		
DT	04-OCT-1996 (Rel. 49, Created)		
DE	05-OCT-1996 (Rel. 49, Last updated, Version 2)		
KW	human STS SHGC-17364.		
OS	Primer; sequence tagged site; STS sequence.		
OC	Human sapiens (human)		
OC	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;		
OC	Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominiidae.		
RP	[1]		
RA	1-241		
RT	Myers R.M.;		
RL	Unpublished.		
CC	Contact: Richard M. Myers, Stanford Human Genome Center (SHGC)		
CC	Stanford University School of Medicine Department of Genetics,		
CC	M-344, Stanford, CA 94305, USA Tel: 4157259687 Fax: 4157259689		
CC	Email: myers@shgc.stanford.edu Primer A: AAGGCTGGATTAAGGAAGACC		
CC	Primer B: GCTGCACCTGGAATGATGC STS size: 150 PCR Profile: Initial		
CC	incubation: 94 degrees C for 90 seconds denaturation: 94 degrees C		
CC	for 15 seconds Annealing: 62 degrees C for 23 seconds		
CC	Polymerization: 72 degrees C for 30 seconds PCR Cycles: 30 Thermal		
CC	Cycler: Perkin Elmer 9600 Protocol: Template: 25 ng Primer: each 1		
CC	ul dNTPs: each 200 uM Taq Polymerase: 0.05 units/ul Total Vol: 10		
CC	ul Buffer: MgCl2: 2.5 mM KCl: 50 mM Tris-HCl: 20 mM pH: 8.3		
CC	Prepared with primer pairs derived from T94579--Merck/UniEST.		
FT	Key	Location/Qualifiers	
FT	source	1..241	
FT		/organism="Homo sapiens"	
FT		/map="1"	
FT	STS	91..240	
FT	primer_bind	91..112	
FT	primer_bind	complement(221..240)	
SO	Sequence 241 BP: 78 A; 53 C; 77 G; 30 T; 3 other:		
Query Match	13.0%; Score 215; DB 250; Length 241;		
Best Local Similarity	97.0%; Pred. No. 0.00e+00;		
Matches	230; Conservative 0; Mismatches 5; Indels 2; Gaps 2		
Db	7	gangggcacaacccaagatttatttttcgaagtgaagggcgacc-aggagaccagaa	65
Cp	1646	GAGGGGCAACCAACCAAGATTATTATTTCAGATGAAGGGGCGAGCCCGAGACCCAGAA	1587
Db	66	aaaagaagagcgaaagctagagcagaagaagcctgtgtaagaagaagacacaaagaagcctg	125
Cp	1586	AAAAGAGAGCGAAGGCTGAGAGACAAAGAGCCCTGGATTAAAGAGACCAAGAAAGGCGCTG	1527
Db	126	cagagcgcaagatttgsgcgccccaaggaagaaacccaagagcgagctgttagtgatctctg	185
Cp	1526	CAGGAGCCAGATGTCGGGCCCCCAGGAAACCCAGAGAGCAGGCGTGTGATGATCTCGG	1467
Db	186	ggcccgacctcaagcttggaacttggaggggctttan-gactcaattcagtgtagca	241
Cp	1466	GCCAGCCTCAAAAGCTGGGACTGGAGCTGAGGAGGCGCTTTTACGCACTCAATTCCAGGTCAGACA	1410
RESULT	4	G29124	241 BP
LOCUS	DEFINITION	human STS SHGC-17364.	DNA
ACCESSION		G29124	STS
			04-OCT-1996

NID 91592433  
 STS sequence: primer; sequence tagged site.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 241)  
 AUTHORS Myers,R.M.  
 JOURNAL Unpublished (1996)  
 COMMENT  
 Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myerseshgc.stanford.edu  
 Primer A: AAGCCTGATTAAGAGAGAC  
 Primer B: GCTGCACCTGGAATTGAGTC  
 STS size: 150  
 PCR Profile:  
 Initial incubation: 94 degrees C for 90 seconds  
 Denaturation: 94 degrees C for 15 seconds  
 Annealing: 62 degrees C for 23 seconds  
 Polymerization: 72 degrees C for 30 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600  
 Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Tag Polymerase: 0.05 units/uL  
 Total Vol: 10 uL  
 Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 20 mM  
 pH: 8.3  
 FEATURES  
 source Prepared with primer pairs derived from T94579--Merck/UniEST.  
 location/Qualifiers  
 1..241  
 /organism="Homo sapiens"  
 /map="1"  
 STS  
 primer\_bind 91..240  
 91..112  
 primer\_bind complement(221..240)  
 BASE COUNT 78 a 53 c 77 g 30 t 3 others  
 ORIGIN  
 Query Match 13.0%; Score 215; DB 202; Length 241;  
 Best Local Similarity 97.0%; Pred. No. 0.00e+00;  
 Matches 230; Conservative 0; Mismatches 5; Indels 2; Gaps 2  
 Db 7 gangggcacaacccaagattattttgccaagttaaggggagcc-aggagaccagaa 65  
 Cp 1646 GAGGGGCACAAACCAAAATTATTATTGGCAAGTAAAGGGGCACCCAGAGACCCAGAA 1587  
 Db 66 aaaaggaaggaagagctagaagcagaagccttgataaagaagacacagaagagcctg 125  
 Cp 1586 AAAAGGAGGCGCAAGGCTGAGCGCGAANAAGCCTGATTAAGGAGAACACACAGAAAGCCTG 1527  
 Db 126 caggaagccagatttcgagccccaagggaacccaagcagagcgtgtagaagtgaacctg 185  
 Cp 1526 CAGGAGCCAGATTGGGGCCCCCAGGAGAAACCCAGAGGAGCGAGGCTGTAGAGTATCTCG 1467  
 Db 186 gccagagctcaaaagcttgacttgtaaggggctttn-gactcaattcagtgagca 241  
 Cp 1466 GCCAGCCTCAAAAGCTGGAGCTGAGGGGGCTTTAGCAGACTCAATTCAGGTGACAGCA 1410

RESULT	5	H43270	405 bp	mRNA	EST	31-JUL-1995
LOCUS		YP05812.1	Homo sapiens cDNA clone 186526 5'		similar to gb:M80927	
DEFINITION		CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.				
ACCESSION		H43270				
NID		9919322				
KEYWORDS		EST.				
SOURCE		human clone-186526 library-Scores breast 3MdBst vector-PT73D (Pharmacia) with a modified polylinker host-DHI0B (ampicillin resistant) primer-M13Rpl Raitel-Not I Raitel2-Eco RI adult human. 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5' TGTTACCAGTCGATCGAGGAGCCGCCCTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified PT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonadata; Neuriopoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 405)				
AUTHORS		Hillier,L., Clark,N., Dubnue,T., Elliston,K., Haskins,M., Holman,M., Holtman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaastis,E., Waterson,R., Williamson,A., Wohlmann,P., and Wilson,R.				
TITLE		The WashU-Merk EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Wilson RK WashU-Merk EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 323 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1..405				
FEATURES		/organism="Homo sapiens"				
source		/clone="186526"				
BASE COUNT		106 a 112 c 95 g 86 t 6 others				
ORIGIN						
Query Match		4.8%; Score 79; DB 41; Length 405;				
Best Local Similarity		63.0%; Pred. No. 1,44e-100;				
Matches		196; Conservative 0; Mismatches 114; Indels 1; Gaps 1;				
Db	1	gctgtcctccagatgccttggacgcgttcnctgtaccacacatcatcaagctttgcca 60				
Oy	145	gcttctcgcccacaaggacttggaacctgacacctgttgacacctcattgacgcttg 204				
Db	61	ataagaagaagatatcacatcgacacctgtgagtagtaagtatgatgacgtctctacggcatagc 120				
Oy	205	gcatacaccacacacacagctgacgccaccactcgaattggaattacacagactctcttcacagaa 264				
Db	121	tcaaacacactcaagaacaggaaccccccaacctgaagactctctgtcgtgaggaagtga 180				
Oy	265	tcaaatggcgctgaaagaagatgattcccaagcttaaaacacctgttaagccatcgagagctgga 324				
Db	181	nenttggtgtctaaagatttccaaagatagcctccaacacccagagtcgcgagacttta 240				
Oy	325	atttcagacactcagaagttcacagattatgtagtgcacacggccaaacacacctcagacctttg 384				
Db	241	tcaagtagtaagtcacatcttggcgcaaccatagcttgatattggtggtctgagacttgcctgg 300				
Oy	385	tcaacttcggccatcagcgtttctgcgcaaatatcagcgttttacggc-ccttgaccttgactgg 443				

D6	301	gtctaccctg	311
OY	444	GAGTACCACG	454
RESULT	6	N75653	EST
LOCUS		541 bp	mRNA
DEFINITION		HY29601.f1 Homo sapiens cDNA clone 244128 5'	similar to gb:M80922
ACCESSION		CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.	
NID		N75653	
KEYWORDS		q1238231	
SOURCE		EST.	
ORGANISM		human clone=244128 primer-reverse ET library=Soares fetal liver spleen INFLS vector-p1713D (Pharmacia) with a modified polylinker host-PHI08 (ampicillin resistant) Rsite1-Pac I Rsite2-Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(ct) primer [5'] AACTGGAAATTAATTAAAGATCTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pvu II and cloned into the Pac I and Eco RI sites of the modified pTV7 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo.	
REFERENCE		Homo sapiens	
AUTHORS		Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes Sarcopterygii; Chonactia; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 541)	
TITLE		Hillier,L., Clark,N., Dubuque,T., Elliston,R., Hawkins,M., Holman,J., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,T., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.	
JOURNAL		The WashU-Mercer EST Project	
COMMENT		Unpublished (1995)	
FEATURES		Contact: Wilson RK WashU-Mercer EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 441 Source: IMAGE Consortium, LMLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Organism 1..541 /organism="Homo sapiens" /clone="244128"	
BASE COUNT		149 a 134 c 117 g 136 t 5 others	
ORIGIN		<1...541	
Query Match		4.7% Score 78; DB 113; Length 541;	
Best Local Similarity		66.7%; Pred.No. 1.34e-98;	
Matches		166; Conservative 0; Mismatches 82; Indels 1; Gaps 1.	
D6	27	ctgggcaggtgtagtgcgtcgtcttcctccagaaggagatcgctcaacaagtgttg	86
OY	44	CTGGGCAGGTTCATGTCGCCGCGTGANTGATCCCATGGGGCCTGCTGCCAACAAGGCTG	103
D6	87	ctactttccaactcgtgtccagacgccggcagaaacaaa-ttcaacctgatgaat	145
OY	104	CTACTTCAACCAACGGGCCACAGTACAGCACAGGGGGAGCTGCTCCGCCCAGAAGACTT	163
D6	146	tgaacctctcatgctctcatcatcatctatcttccttgcagagatgtaaaaaaacagyt	205
OY	164	GGAACCCACAGCCTTTGACACCCACCTCATCTTAGCCTTGCGTGGCATGACCAACACCACT	223
D6	206	tattcataaggacaagatgaagtatgctctaccagacatacaaacagctctcaaaccaa	265

OY 224 GAGCACCACGAGTGGAAAGACGAGACTCTCTACGAGGATTCATAGGCGCTGAAGAAGAT 283  
 Db 266 gaatcccaa 274  
 OY 284 GAATCCCAA 292

RESULT 7 541 bp mRNA EST 02-APR-1996  
 LOCUS N75653  
 DEFINITION yv29c01.r1 Homo sapiens CDNA clone 244128 5' similar to gb:M80927  
 CARRIER GLYCOPROTEIN-39 PRECURSOR (HUMAN);.  
 ACCESSION N75653  
 NID 91238231  
 KEYWORDS EST.  
 SOURCE human clone-244128 primer-reverse ET library-Soares fetal liver  
 spleen INFIL vector-p773D (Pharmacia) with a modified polylinker  
 host-DH10B (ampicillin resistant) Rstlet-Pac I Rstlet-Eco RI liver  
 and spleen from a 20 week-post conception male fetus. 1st strand  
 CDNA was primed with a Pac I - oligo(dT) primer [5']  
 AACGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3', double-stranded  
 CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac  
 I and cloned into the Pac I and Eco RI sites of the modified p773  
 vector. Library went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 541)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 High quality sequence stops: 441  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES  
 source NCBI g1: 1238231  
 1..541 Location/Qualifiers  
 /organism="Homo sapiens"  
 /clone="244128"  
 /note="human"  
 mRNA 149 a 134 c 117 g 136 t 5 others  
 BASE COUNT  
 ORIGIN

Query Match 4.7%; Score 78; DB 207; Length 541;  
 Best Local Similarity 66.7%; Pred. No. 1.34e-98;  
 Matches 166; Conservative 0; Mismatches 82; Indels 1; Gaps 1;  
 Db 27 ctggagaggtgtatgtctgtctgtctctcagaggagagctgcctacaactgtgttg 86  
 OY 44 CTGGAGAGGTTTATGTCTGTCTGTCTGATGATCCATGGGCTCTGCTGCAAAATGCTG 103  
 Db 87 ctacttacaactgtgtccagagacggcaggaacaggaana-ttaaccttgaagaat 145  
 OY 104 CTACTTCACCACTGGCCAGTACAGACAGGGGAGGCTCGCTCTGCCCAAGACTT 163

Db 146 tgacccttcctatgtctctcatcatcatctatctatccgcagcatcgaacaaacaggt 205  
 OY 164 GGACCCACGCTTTTGCACCCACTCATCTACGCTTTTGGTGGCATGACCAACCAAGCT 223  
 Db 206 tatcatcaagagcaagagtgatgtctctaccagacatcaacagctcaaaccaa 265  
 OY 224 GAGCACCAGTGAAGTGAAGAGAGAGACTCTACACGAGGATTCATAGGCGCTGAAGAAGAT 283  
 Db 266 gaatcccaa 274  
 OY 284 GAATCCCAA 292

RESULT 8 533 bp mRNA EST 20-FEB-1995  
 LOCUS yv24b03.s1 Homo sapiens CDNA clone 81581 3' similar to sp:S27879  
 DEFINITION S27879 SECRETORY PROTEIN YM-1 PRECURSOR - .  
 ACCESSION T65854  
 NID 9674899  
 KEYWORDS EST.  
 SOURCE human clone-81581 library-Stratagene lung (#937210)  
 vector-pBluescript SK- host-SOLR cells (kanamycin resistant)  
 primer-23ml3 Rstlet-EcoRI Rstlet-XhoI Normal lung tissue from a 72  
 year old male. Cloned unidirectionally. Primer: Oligo dT. Average  
 insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:  
 5'-GATTCGGCAGCAG-3'; 3' adaptor sequence:  
 5'-CTCGAGTTTCTTTTCTTTT-3'.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 533)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rikkin, L., Rohlfing, T., Tan, F., Trevasakis, E.,  
 Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 High quality sequence stops: 307  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES  
 source 1..533  
 /organism="Homo sapiens"  
 /clone="81581"  
 BASE COUNT 140 a 123 c 136 g 128 t 6 others  
 ORIGIN

Query Match 4.4%; Score 73; DB 144; Length 533;  
 Best Local Similarity 70.1%; Pred. No. 7.80e-89;  
 Matches 131; Conservative 0; Mismatches 55; Indels 1; Gaps 1;  
 Db 250 tggacaagagggttaag 309  
 Cp 1162 TGGATGAGGGGGATTCGGCCGTGGTGCAGAGAACCGGCAAGATCATTAATTCACCT 1103  
 Db 310 gaccagatcatgtctctccagaggttaaatctttaagaactgaactgtgtctccata 369  
 Cp 1102 GCCAGACCATGGCCCCGCCAGTCCCTCTCTGCTTCAGATACCTGCTTTGAAG 1043  
 Db 370 ctcttcaatcatatagaccacacacacacacacacacacacacacacacacacacac 429  
 Cp 1042 CTCCTCCATCATCAAGGCCACCACT-GGTTGCCGGAAGAGATGAGGACACCTTCTG 984  
 Db 430 atctcttg 436

CP 983 ATCTGG 977

|||||

RESULT 9 G24226 533 bp DNA STS 31-MAY-1996

LOCUS G24226

DEFINITION human STS WI-14165.

ACCESSION G24226

NID G1344552

KEYWORDS STS sequence; primer; sequence tagged site.

SOURCE human STS derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 533)

AUTHORS Hudson, T.

TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS

JOURNAL Unpublished (1995)

COMMENT

Contact: Thomas Hudson

Whitehead Institute/MIT Center for Genome Research

Whitehead Institute for Biomedical Research

9 Cambridge Center, Cambridge MA 02142 USA

Tel: 617 252 1900

Fax: 617 252 1902

Email: thudson@genome.wi.mit.edu

Primer A: AGGCGACAGATGACACGCT

Primer B: TCTCCAGATTCATGCTGG

STS size: 132

PCR Profile:

Presoak:

Denaturation: Annealing: 56 degrees C

Polymerization: PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pm

dNTPs: each 4 mM

Taq Polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCL: 10 mM

pH: 9.3

Derived from dbEST (genbank accession T65854).

FEATURES

source

1..533

Location/Qualifiers

/organism="Homo sapiens"

/map="461.5 CR from top of Chr1 linkage group"

STS

primer\_bind 34..165

34..53

complement(145..165)

BASE COUNT 140 a 123 c 136 g 128 t 6 others

ORIGIN

Query Match 4.4%; Score 73; DB 200; Length 533;

Best Local Similarity 70.1%; Pred. No. 7.80e-89;

Matches 131; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

Db 250 tggacaagaggttaagagccctggttcaggaattncacagtgacatccatgcaata 309

|||||

CP 1162 TGGATGAGGGGATCGGCGCTGGTTGCAAGAGCCGCAAGTCATCTAAGTCCAGT 1103

|||||

Db 310 gaccagatcatggtctctccacaggttaaatctttaagaactgaacctggtctccata 369

|||||

CP 1102 GCCCAGACATGCGCCCGCCAGTCCTGCTTCAGATAGTGCATCTGTTTGAAG 1043

Db 370 cttcacatcatatagccaccactgttcccttgcactcgtgaaggagactctg 429

|||||

CP 1042 CTCTCACATCATCAAGCCACCACCT-GGTTGTCCCGAAGATGAGGACCTTCTG 984

Db 430 atctctg 436

|||||

CP 983 ATCTGG 977

RESULT 10

LOCUS M40107 311 bp mRNA EST 22-JAN-1996

DEFINITION yv72b09.r1 Homo sapiens cDNA clone 257753 5' similar to gb:M80927

ACCESSION CARIAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.

NID M40107

KEYWORDS 91163652

SOURCE EST.

human clone-257753 primer-T7 library=Soares placenta 8to9weeks

2NbHpt09M vector-PT7T3D (Pharmacia) with a modified polylinker

host-DH10B (ampicillin resistant) Rstrel-Not I Rstle2-Eco RI two

placentae: one from 8 weeks and another from 9 weeks post

conception. 1st strand cDNA was primed with a Not I - oligo(dt)

primer [5'-TGTACCATCTGAGTGGAGCGCGCGATTTTCTTTTCTTTT-3'],

double-stranded cDNA was size selected, ligated to Eco RI adapters

(Pharmacia), digested with Not I and cloned into the Not I and Eco

RI sites of a modified pT73 vector (Pharmacia). Library

constructed by Bento Soares and M. Fatima Bonaldi.

ORGANISM Homo sapiens

Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 311)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The Washu-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK

Washu-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 268

Source: IMAGE Consortium, LBNL

This clone is available royalty-free through LBNL; contact the

IMAGE Consortium (info@image.lbnl.gov) for further information.

FEATURES

source

1..311

Location/Qualifiers

/organism="Homo sapiens"

/clone="257753"

mRNA

<1...>311

BASE COUNT 93 a 81 c 68 g 68 t 1 others

ORIGIN

Query Match 4.3%; Score 71; DB 79; Length 311;

Best Local Similarity 65.8%; Pred. No. 5.83e-85;

Matches 156; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

Db 69 ctggcaggtttagtgcgtctctccaggaagatcgtccataactggttg 128

|||||

QY 44 CTGGCAGATTTCATGCGTCGCGATGATGATCCATGGGCGCTCTCTGCAAAACGTCGTG 103

|||||

Db 129 ctacttaccacatggtctccaggaacggcaggaacaggaaat-ttcaacctggaatat 187

|||||

QY 104 CTACTTCAACCACTGGGCGCATGACAGACAGGGGAGGCTGCTCTGCCAAGGACTT 163

|||||



Db	188	tgaaccttcctcatgtctcctcatcctatccattccatgcgacagatcgaaacaagaagt	247
Oy	164	GGACCCAGACCCTTTGACACCACTTATCTACAGCCTTGCGTCGCATGACACCAACAAGCT	223
Db	248	tatcataagagcaagaagtgaagtgtatgtctctacagaccatacaagctctcaaaaa	304
Oy	224	GAGCACCAACTGTGGATGGAATGACGAGAGCTCTCTACGACGAGATTCAATGGCCTGAAGA	280

RESULT	11	
LOCUS	W51302	336 bp mRNA EST 29-MAY-1996
DEFINITION	ma16f08.r1 life Tech mouse brain Mus musculus cDNA clone 304743 5'	
ACCESSION	similar to gb:X93035 M.musculus mRNA for BRP39 protein (MOUSE).	
NID	W51302	
KEYWORDS	g13397/42	
SOURCE	EST.	
ORGANISM	house mouse. Mus musculus	
REFERENCE	Eukaryotae: mitochondria eukaryotes; Metazoa: Chordata; Vertebrata: Eutheria: Rodentia: Sciurognathi: Myomorpha: Muridae; Murinae: Mus. 1 (bases 1 to 336)	
AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubaque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	
TITLE	The Mashu-HHMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: musese@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMGE Consortium (info@image.jnl.gov) for further information. MG1:201503 Seq primer: ETPRimer High quality sequence stop: 148. Location/Qualifiers 1..336 /organism="Mus musculus" /strain="C57BL/6J" /note="Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn. pCMV-SPORT2 vector." /clone_id="304743" /clone_lib="Life Tech mouse brain" /dev_stage="adult" /lab_host="DH10B" <1...>336	

FEATURES	
SOURCE	
ORIGIN	
BASE COUNT	83 a 100 c 89 g 64 t
MRNA	
ORIGIN	

Query Match	4.3%; Score 71; DB 169; Length 336;
Best Local Similarity	69.2%; Freq. No. 5.83e-85;
Matches 137; Conservativity	0; Mismatches 60; Indels 1; Gaps 1;

Db	140	caggcttgcagctcctcgaatgcgtctcacagagctgcgtcgtcgatacaagtgtgtctact	199
Oy	49	CAGGTTTCAATGCTCTCTCGATGATGATCCATGGGCTCTGCTGCAAAAGTGCTGTACT	108
Db	200	acaccaactggtcccaagtaaccggaaaggcaatyggagctgtctccagagacgccctgacc	239
Oy	109	TCACCAACTGGGCCAGTACAGACAGGGGGAGGCTGCTTCCTCCGCCAAGGACTTGAGAC	168
Db	260	attcccttgtagccccaatatcatctaagtttggcaa-catcagcacaacaagcttagca	318
Oy	169	CCAGGCTTTGCAACCACTCATATCTTAGGCTTCCCTGCGATGACCAACACCAAGCTGAGCA	228

Db	319	catccgagtggaatgcacg	336
OY	229	CCACTGAGTGGGAATGCACG	246
RESULT	12	H24821	266 bp mRNA EST 07-JUL-1995
LOCUS		Y142F01.r1 Homo sapiens cDNA clone 160921 5'	similar to gb:M809227
DEFINITION		CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.	
ACCESSION		H24821	
NID		9893720	
KEYWORDS		EST.	
SOURCE		human clone-160921 library-Scares breast 3NHBst vector-pt773D (Pharmacia) with a modified polylinker host-pHI0B (ampicillin resistant) primer-M13RP1 Rsite1-Not I Rsite2-Eco RI Adult human. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTCACCATCGAAGTGAGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of modified pRT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fátima Bonaldo.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrate; Gnathostomata; Osteichthyes; Sarcopterygii; Channata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 266)	
AUTHORS		Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.	
TITLE		The WashU-Merk EST Project	
JOURNAL		Unpublished (1995)	
COMMENT		Contact: Wilson RK WashU-Merk EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 84 Source: IMAGE Consortium, LNL. This clone is available royalty-free through LNL : contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Location/Qualifiers 1..266 /organism="Homo sapiens" /clone="160921"	
FEATURES		BASE COUNT	58 a 80 c 70 g 51 t 7 others
ORIGIN		Query Match	4.0%, Score 66; DB 35; Length 266; Best Local Similarity 68.3%; Pred. No. 2.27e-75; Matches 127; Conservative 0; Mismatches 58; Indels 1; Gaps 1,
Db	14	ggaccacaagtctcataagaatcctctgtggcaggcaggtcccttatgccaccaaggcgcaaccagt	73
OY	958	GGGCGACCACCAACAAGAATCAGATACAGAAAGGTGCCCTCACATTCTCGGACCAACAGT	1017
Db	74	gggttagataacagacagaccagaagagctcaaaagcacagtcgacgtacctgaagaaagcgc	133
OY	1018	GGGTGGGCTTTGATGATGATGTGGAGAGCTTCCAACCAAGGTCAGCTATCTTAAGACAAAGG	1077
Db	134	agcttgcg-cgggcgatgtgatatggccccttnaacctgtgatnactincagggtctcttcgcg	192
OY	1078	GACTGGGGGGGGCCATGCTGTGGCACTGAGACTTAGATGACTTTGGCGGCTTGCTCGCA	1137
Db	193	gcgcagc 198	
OY	1138	ACCAGG 1143	

RESULT	13				EST	22-MAY-1995		
LOCUS	R55530	396 bp	mRNA					
DEFINITION	y729c12.r1 Homo sapiens CDNA clone 154966 5' similar to gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.							
ACCESSION	R55530							
NID	9824825							
SOURCE	EST. human clone=154966 library-soybean breast 2NBHST vector-pt73D (pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) primer-M13RPI Rsite1-Not I Rsite2-Eco RI Adult female 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCACCACTCGAAGTGAGGAGCCGCCGCCTTTTATTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of modified pTR73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo.							
ORGANISM	Homo sapiens Eukaryote; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 396) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.							
TITLE	The Wash-Merck EST Project							
JOURNAL	Unpublished (1995)							
COMMENT	Contact: Wilson RK Wash-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 284 Source: IMAGE Consortium, LNUH This clone is available royalty-free through LNUH : contact the IMAGE Consortium ( <a href="mailto:info@image.lnuh.gov">info@image.lnuh.gov</a> ) for further information. Location/Qualifiers 1..396 /organism="Homo sapiens" /clone="154966"							
BASE COUNT	88 a	106 c	109 g	87 t	6 others			
ORIGIN								
FEATURES								
source								
Query Match	3.9%	Score 65;	DB 134;	Length 396;				
Best Local Similarity	68.3%;	Pred. No. 1,81e-73;						
Matches 114;	Conservative 0;	Mismatches 53;	Indels 0;	Gaps 0;				
Db 125	cagccttgatgtgcctgtgtgcctgcacgaibcctctcgatatacaactgctgtact	184						
OY 49	CAGGTTTCATGTGCTCTGCTGTATGATATCCATGGGGCTGTGCTGCACAAACTGATCTGACT	108						
Db 185	aacaccagtgtccacgtaccgganangcgtatggagctgcttcaccagatgccttgaac	244						
OY 109	TCACCAACTGGGCGCAGACAGACAGGAGGAGGCTCGCTTCCTGCCCAAGACTTGGACC	168						
Db 245	gnntccctctgacccacatatcatcacagctttgccaatataagaac	291						
OY 169	CCAGCCTTTGCACCCACACTCATCTACAGCCTTCGCTGGCACATGACACAC	215						
RESULT	14							
LOCUS	w10705	306 bp	mRNA	EST	26-APR-1996			
DEFINITION	ma47907.r1 Soares mouse p3NM19.5 Mus musculus CDNA 5' similar to gb:M80927 CARTILAGE GLYCOPROTEIN-39 PRECURSOR (HUMAN);.							
ACCESSION	w10705							

NID	g1285179	EST.	house mouse.	
KEYWORDS	mus musculus			
SOURCE	Eukaryote; mitochondrion			
ORGANISM	Eukaryote; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 306)			
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Treisman, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.			
TITLE	The WashU-HMNI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra, M/Mouse EST Project WashU-HMNI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: mob.REGA+ET High quality sequence stop: 172.			
FEATURES	NCBI gi: 1285179			
SOURCE	Location/Qualifiers 1..306 /organism="Mus musculus" /note="vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACATCTGACATGAGGAGCGGCGCATTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo." /clone_1lb="Soares mouse p3NM19.5" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)" <1..>306			
mrna	78 a	76 c	85 g	67 t
BASE COUNT				
ORIGIN				
Query Match	3.6%;	Score 60;	DB 215;	Length 306;
Best Local Similarity	72.1%;	Pred. No. 4.65e-64;		
Matches	98;	Conservative 0;	Mismatched 38;	Indels 0;
Gaps	0;			
Db	78	agaagcttccttcgctaccacgaaggaacacagctgggtggtatgtagagacaagagatg	137	
QY	985	AGAAggtccctTACATCTTCGCGGACAAACGATGGGTGGCTTTGATGATGTGGAGACT	1044	
Db	138	tcaaaaacaaggttgggtctctgaaggaagaaagctggcagagcatcggttctggcac	197	
QY	1045	TCAAAACCAACGATCAAGTCTGTAAGCAGAGAGGACTGGCGGCCATGTGTGGGCAC	1104	
Db	198	tggtattgtagatctt	213	
QY	1105	TGGACTTAGATGACTT	1120	
LOCUS	15			
LOCUS	W10705	306 bp	mrna	EST
DEFINITION	ma47g97.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone 313884 5'			
ACCESSION	W10705			
NID	g1285179			
KEYWORDS	EST.			

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Mus.  
REFERENCE 1 (bases 1 to 306)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: mob.REGA+ET  
High quality sequence stop: 172.  
Location/Qualifiers  
1..306  
/organism="Mus musculus"  
/note="Vector: p773D (Pharmacia) with a modified  
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TCTTCCACTCTGAGATGGAGCGCGCCGATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Falima Bernaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."  
/clone="313884"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
<1..>306  
MRNA  
BASE COUNT 78 a 76 c 85 g 67 t  
ORIGIN  
Query Match 3.68; Score 60; DB 174; Length 306;  
Best Local Similarity 72.18; Pred. No. 4.65e-64;  
Matches 98; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
Db 78 agaaggtcccttcgctcgaagagcaccagctgggtggtatgaggacaagagatg 137  
||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 985 AGAAGTCCCTTCATCTTCGCGACACCACTGGGTGGCTTGATGATGTGACAGCT 1044  
Db 138 tcaaaacaaggttgggtcctgaagagaagaagctggcaggacatcgttggcac 197  
||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1045 TCAAAACAAGGTCTCATCTCTCAAGCAGAAAGGACTGGCGGCCCATGCTCGGCGAC 1104  
Db 198 tggatttgatgattt 213  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1105 TGGACTTAGTGACTT 1120

Search completed: Sun Jun 29 21:08:58 1997  
Job time : 1195 secs.







PD 30-JAN-1988.  
 PF 16-JUL-1986; 167518.  
 PR 16-JUL-1986; JP-167518.  
 PA (DAIN) Daiinippon Pharm KK.  
 DR WPI: 88-068419/10.  
 P-PSDB: P81342.  
 PT New polypeptide and DNA encoding it - related to protective mechanisms such as immune response etc.  
 PS Disclosure; p; Japanese.  
 CC This polypeptide is involved in protective mechanisms such as immune response, cell growth and activation of protective functions.  
 SQ Sequence 321 AA;

Query Match 26.3%; Score 906; DB 1; Length 321;  
 Best Local Similarity 49.1%; Pred. No. 2,43e-72;  
 Matches 113; Conservative 66; Mismatches 45; Indels 6; Gaps 3;

Db 7 tgfvtvlllgccsasklvcytwsgyregdscfdaldfflctchlysfansndhid 66  
 QY 8 AGFVLMIMPWSAKLVCTFTNNAQYRGEARFLPKLDSLCYHLIYARAGMTNHL 67  
 Db 67 twewndvlygmlltknnpulklisvsgwmfsgqrfsklasntgstrtfkissvpl 126  
 QY 68 TTEWNDVLYGMFTLTKNRPULKLISVSGWMFSGQRFKLSANTGSTRTFKISSVPL 127  
 Db 127 rthfgdvudlawlypg-rg----dkqhttltkemkaefieag-pqkqqlilsaalsag 180  
 QY 128 RKFYFDDLDLWEPYSGSGSPAVDKERTTLVQDLANAFQQAQSGERLLLSAAYVAG 187  
 Db 181 kvtdssydlaklsqhlidlsimtydfhgwargtqthspilfrggedasp 230  
 QY 188 QTYVDAGLEVDKINQNDLVNLMAYDFHGSWEKVTGHSPLXKQESGA 237

RESULT 5  
 ID R70025 standard; Protein; 371 AA.  
 AC R70025;  
 DT 29-SEP-1995 (first entry)  
 DE Tobacco chitinase encoded by tobacco Cluster-A cDNA clone CA-3.  
 KW Tobacco; chitinase; antifungal; fungicide; Cluster-A.  
 OS Nicotiana tabacum.  
 FH Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /label= signal peptide  
 FT MISC\_difference 22  
 FT /note= "not present in genomic clone"  
 FT MISC\_difference 182..195  
 FT /note= "Determined by protein sequencing"  
 FT MISC\_difference 216..241  
 FT /note= "Determined by protein sequencing"  
 PN EP-639642-A.  
 PD 22-FEB-1995.  
 PF 17-AUG-1993; 202425.  
 PR 17-AUG-1993; EP-202425.  
 PA (MOGE-) MOGEN INT NV.  
 PA (OTLE-) RIJUSUNTY LEIDEN.  
 PI Apoltheker-de Groot M, Bol JF, Cornelissen BJC, Linthorst HJM;  
 PI Melchers IS, Ponstein AS, Sela-baurlage MB;  
 DR WPI: 95-083454/12.  
 DR N-PSDB: 082973.  
 PT New plant protein having endo-chitinase activity - used in antifungal compns. and to develop transformed plants which are less susceptible to fungal infection.  
 PS Claim 3; Page 19-20; 43pp; English.  
 CC Screening of a lambda ZAP cDNA library of TMV-infected Samsun NN tobacco plants with a probe derived from PROB40 (a partial Cluster-A cDNA clone) resulted in the isolated of 11 positively hybridizing clones. Analysis revealed that all were identical and corresp. to Cluster-A cDNA. The nt sequence of clone CA-3 is given in 082973/ R70025. This clone lacks 7 codons of the 5' part of the ORF. The cDNA clone was completed by performing a PCR reaction on clone CA-3 with primer T7 (082974) and primer P1 (082975). A partial AA sequence of the protein isolated from TMV-infected tobacco leaves

CC was obt'd. and found to be almost identical to the AA sequence  
 CC deduced from the cDNA.  
 SQ Sequence 371 AA;

Query Match 17.4%; Score 597; DB 13; Length 371;  
 Best Local Similarity 33.0%; Pred. No. 3.72e-43;  
 Matches 107; Conservative 86; Mismatches 108; Indels 23; Gaps 18;

Db 36 nidsstflthlfcadlnpsngllspengds-fsqfstvgrknpvskffisaggr 94  
 QY 44 KDLPSICTHLIYAFAGMT--NHOLSTEWDELYQEFNG-LKKMNPKLTLLAI-GGW 99  
 Db 95 a-dftaylmargpsrksfidsrlargqfngldweyp--l-s-atdmnlglll 149  
 QY 100 NFSTQKFTDMVATNANNQTFVNSAIRFLRYSFVGLDLDWEPYSGSGSPADKRFETLV 159  
 Db 150 newrtalmeearnsgraailltaavysprynglnyvesvarlnvlnmaydfygnw 209  
 QY 160 QDLNNAQQAQSGTSGKERLLLSAAYVAGQTYVDAGYEVDKRAQNDLVNLMAYDFHGS-W 218  
 Db 210 spsgtushaglfddvnh--isgs--dg-lnawlgayvptkklvlgpfgyawrvn 263  
 QY 219 EKV-TGHSNPLYKROESGAASLNVDAAVQWQKQGPASKLIGMPTYGRSFTLIASS 277  
 Db 264 lhdrlapaagksnyga-vdggsm-tynrirdyvtqratvynatlygdvysgsmly 321  
 QY 278 DTRVAGAPATGSGTPGPTKREGMLAYEVCSWKQATKOR-IQDOK-VP-YIFRDNQWGF 334  
 Db 322 ddtgsvrnknyvkgfqllyfaw 345  
 QY 335 DDVESFKTKVSYLKQKIGGAMW 358

RESULT 6  
 ID R70029 standard; Protein; 377 AA.  
 AC R70029;  
 DT 29-SEP-1995 (first entry)  
 DE Tobacco cluster-A protein encoded by genomic clone.  
 KW Tobacco; chitinase; antifungal; fungicide; Cluster-A.  
 OS Nicotiana tabacum.  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /label= signal peptide  
 FT MISC\_difference 28..29  
 FT /note= "cDNA encodes additional Ile here"  
 FT MISC\_difference 188..201  
 FT /label= Determined by sequencing the protein  
 FT MISC\_difference 222..247  
 FT /note= "Determined by sequencing the protein"  
 PN EP-639642-A.  
 PD 22-FEB-1995.  
 PF 17-AUG-1993; 202425.  
 PR 17-AUG-1993; EP-202425.  
 PA (MOGE-) MOGEN INT NV.  
 PA (OTLE-) RIJUSUNTY LEIDEN.  
 PI Apoltheker-de Groot M, Bol JF, Cornelissen BJC, Linthorst HJM;  
 PI Melchers IS, Ponstein AS, Sela-baurlage MB;  
 DR WPI: 95-083454/12.  
 DR N-PSDB: 082976.  
 PT New plant protein having endo-chitinase activity - used in antifungal compns. and to develop transformed plants which are less susceptible to fungal infection.  
 PS Claim 3; Page 25-28; 43pp; English.  
 CC Screening of a lambda ZAP cDNA library of TMV-infected Samsun NN tobacco plants with a probe derived from PROB40 (a partial Cluster-A cDNA clone) resulted in the isolated of 11 positively hybridizing clones. Analysis revealed that all were identical and corresp. to Cluster-A cDNA. The sequence of cDNA clone CA-3 is given in 082973/ R70025. A genomic library of N. tabacum was screened using the Cluster-A cDNA insert of clone CA-3 as a probe (see 082977, 082978). The complete nt. sequence of Cluster-A cDNA including the deduced primary structure of the Cluster-A protein the 5' and 3' UTR regions of the gene are shown in 082976/R70029. Comparison of the cDNA









QY 227 PLKROESGAASLNDAAVQOOLKGTPTASKLILMPTGRSFTLIASSSDTRVGA PAT 286  
 DB 463 hravkqvw--englvdyrqqlasqfmsqewqytdgtteapvfvfpgsdlttsdarsv 520  
 QY 287 GSGPGEFTTEGGGLAYECS-WKGTQKQ-RIQD-QKVEYFR-DN-QWVGEDVESFK 341  
 DB 521 akgyvldkqjgglfsweldadn 543  
 QY 342 TKVSYLKQKGLGAMWALDDDD 364

RESULT 14  
 ID R32547 standard; Protein: 238 AA.  
 AC R32547;  
 DT 17-JUN-1993 (first entry)  
 DE HANPY IE-1 gene protein (partial sequence).  
 KW Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;  
 OS Heliothis armigera nuclear polyhedrosis virus.  
 PN MO9303144-A.  
 PD 18-FEB-1993.  
 PF 05-AUG-1992; AU0413.  
 PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 PI Christian PD, Dail DJ, Gordon KHU, Hanzlik TN, Srisanthia A.  
 DR MPI: 93-076501/09.  
 DR N-PSDB; Q35982.  
 PT Insect virus with reduced capacity to occlude viral particles - used  
 PT for controlling proliferation of insect pests without horizontal  
 PT transmission  
 PS Disclosure: Fig 2: 51pp: English.  
 CC The sequence is that encoded by the Heliothis armigera (Ha)  
 CC nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene  
 CC (partial sequence).  
 SQ Sequence 238 AA;

Query Match 7.9%; Score 271; DB 6; Length 238;  
 Best Local Similarity 34.1%; Pred. NO. 1.95e-13;

Matches 58; Conservative 43; Mismatches 51; Indels 18; Gaps 14;

DB 44 erdkfinspgwrefvltwkl-fdgldidwefpgkganpugvverndntlyiallgelra 102  
 QY 115 NROTFVNSAI-R-F-L-RKXSPDGLDLDWXPSSQS-PAV-DKER----FTLVQDLAN 164  
 DB 103 mldq-vqlgturtleltalsagldkaav-ncdraqyldkifmsydfkgswnldlg 160  
 QY 165 AFQOEAOISGKERLLSAAPACQTYVDAGYEVDKIQNDPVLMAYPDHSMEKVT-G 223  
 DB 161 hqcaly--gsawkpnepytanvadalagrvnp-kklylgvamygrwt 207  
 QY 224 HNSPLTKROESGAASLNDAAVQOWL-QKGTPTASKLILMPTGRSFT 272

RESULT 15  
 ID W02156 standard; Protein: 1046 AA.  
 AC W02156;  
 DT 14-JAN-1997 (first entry)  
 DE Periplasmic chitodextrinase.  
 KW Periplasmic chitodextrinase; periplasmic Beta-N-acetylglicosaminidase;  
 KW Beta-N-acetylglicosaminidase; chitin, oligosaccharide; catabolic;  
 KW catabolism.  
 OS Vibrio furnissii.  
 PN W09625424-A1.  
 PD 22-AUG-1996.  
 PF 13-FEB-1996; U02332.  
 PA 13-FEB-1995; US-386727.  
 PA (Uxjo ) UNTV JOHNS HOPKINS.  
 PI Bassler B, Chlilaru E, Keyhani N, Roseman S, Rowe C;  
 PI Yu C;  
 DR MPI: 96-393335/39.  
 DR N-PSDB; T36387.  
 PT Chitin biosynthetic enzymes end I, exo I and exo II - are

PT periplasmic chito:dextrinase(s), periplasmic beta-glucosidase(s) and  
 PT aryl beta-N-acetyl:gluco:amidase(s), respectively  
 PS Claim 2: Page 68-71, 101pp; English.  
 CC Periplasmic chitodextrinase (W02156), periplasmic  
 CC beta-N-acetylglicosaminidase (W02157) and aryl  
 CC beta-N-acetylglicosaminidase (W02158) can be used to produce chitin  
 CC oligosaccharides with the structure (GlcNAc)n where n is 2 or  
 CC higher, by contacting them with soluble chitin. The enzymes are  
 CC encoded by the genes endI, exoI and exoII respectively. They are  
 CC all genes involved in the catabolic pathway of chitin.  
 SQ Sequence 1046 AA;

Query Match 6.4%; Score 220; DB 19; Length 1046;  
 Best Local Similarity 31.1%; Pred. NO. 4.80e-09;

Matches 52; Conservative 44; Mismatches 56; Indels 15; Gaps 11;

DB 432 vadqgf-yenttnadgsinqgletfadsavemrkyrfdgldlrinsldgqtgppd 490  
 QY 94 LAIGWNFSQKRTDQVATANNRQTFVNSAIRLRKXSPDGLDLD-W--EYPSQSSPA- 149  
 DB 491 tafserraylmnsyhelmrvtrekldvasagdgqvhymltlaaps-eyllrgmetnavt 549  
 QY 150 VD-KE-RFTLV--ODLANAFQOEAO-TSGKE--RLLSAAVPACQTYVDAGYEVDKIA 201  
 DB 550 qyldvynmsydlhgawndhvgnaalydtgkselagw-nvygtag 595  
 QY 202 QNDPVLMAYPDHSMEKVTGHSPLTKROESGAASLNDAAVQ 248

Search completed: Fri Jun 27 16:59:22 1997  
 Job time : 98 secs.





```
QY 68 TTEHNDLTLQFENGKMKPKLKTLLAIGWNESTQKFTDMVATANNROTFSVSAIRFL 127
Db 128 rthbfddglawlyprrr-----dkhfttltkemkeeflkeag-pqkqjllsaalsag 181
QY 128 RKTSPDLDDMEYPSGSGSPAVDKERFTTLVODLANAFQEOAQTSGKERLLLSAAPAG 187
Db 182 kvldesydaklsqhlidfsimtydfhgwrgtqghsp1ffrqedaspdfrfentdyav 241
QY 188 QTVYDAEYVDKIAQNLDVFNLAAYDFHGSWEKVTGHNSPLYKROESGAASINVDAAV 247
Db 242 gymrljlapasaklvmgipctfgsrftlaas-etyvgapisgpgjprftkkaagtlayyeic 300
QY 248 QQWLQKGTSPASKTLIGMPYGRSFTLLASSSDTRVGAPATGSGPTKEGGMAYEVC 307
Db 301 dflrgatvhtl1ggvgyackngwgyddgeskxvqylktdqqlagammwaldddfq 360
QY 308 SW-KGATKQRIQODQKVPYIRRDNONWGFDDVESFKTKVSTLKQKGLGAGAWMALDDFA 366
Db 361 gsfecgqlrftplnaikdala 381
QY 367 GFSCNQG-RYPLIQTLROELS 386

RESULT 2
ENTRY S51327 #type complete
TITLE heparin-binding glycoprotein 38k - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 07-May-1995 #sequence_revision 07-May-1995 #text_change
ACCESSIONS S51327
REFERENCE S51327
#authors Shackleton, L.M.; Mann, D.M.; Mills, A.J.T.
#submission submitted to the EMBL Data Library, January 1995
#description Identification of a 38kDa heparin-binding glycoprotein (9p38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling.
#accession S51327
#status preliminary
#residues 1-383 ##label SHA
#cross-references EMBL:247803
SUMMARY #length 383 #molecular-weight 42443 #checksum 2907

Query Match 42.2%; Score 1456; DB 16; Length 383;
Best Local Similarity 50.8%; Pred. No. 1,65e-251;
Matches 197; Conservative 88; Mismatches 94; Indels 9; Gaps 5;

Db 1 mg1rvaqtfva1v1qscayk1vcytswsgyregdscfpda1dpfcthl1ysfan 60
QY 1 MVRSVAMAGFVLMIPWGSAAKLVCFETWMAQYRQGEARFLPKDLPISCTHLIYAFAG 60
Db 61 tsmneitlwmndctlydntlnknrpn1k1lsvggwfgsgfrfkiaentsrtffl 120
QY 61 MTNHQLSTTWMNDTLQFENGKMKPKLKTLLAIGWNESTQKFTDMVATANNROTFFV 120
Db 121 ksvppftrthbfddglawlyprrr-----dkhfttltkemkeefreal-pgte1ll 174
QY 121 NSAIRFLKRXSPDLDDMEYPSGSGSPAVDKERFTTLVODLANAFQEOAQTSGKERLL 180
Db 175 sgavsagkvaldrgyda1sqh1d1s1llydfhgwrgtqghsp1ffrqgqdasdrf 234
QY 181 SAAVPAGQTVYDAEYVDKIAQNLDVFNLAAYDFHGSWEKVTGHNSPLYKROESGAAS 240
Db 235 snadyevsylv1jlapan1k1vmgipctfgsrftlaas-ktdgapaspgjprftkky1 293
QY 241 LNWDAAVQWMLQKGTSPASKTLIGMPYGRSFTLLASSSDTRVGAPATGSGPTKEGGM 300
Db 294 layyedf1qgatrprp1ggvgyackngwgyddgeskxkay1lksrq1agammwt 353
QY 301 LAIYEVCSW-KGATKQRIQODQKVPYIRRDNONWGFDDVESFKTKVSTLKQKGLGAGAWMA 359
Db 354 ldllddfgq-tcqpkef1plnaikdala 381
```

```
QY 360 LDLDDEPAGFSCNQG-RYPLIQTLROELS 386

RESULT 3
ENTRY S61550 #type complete
TITLE BRP39 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
ACCESSIONS S61550
REFERENCE S61550
#authors Morrison, B.W.
#submission submitted to the EMBL Data Library, November 1995
#accession S61550
#status preliminary
#molecule_type mRNA
#residues 1-381 ##label MOR
#cross-references EMBL:X93035
SUMMARY #length 381 #molecular-weight 43001 #checksum 5107

Query Match 42.2%; Score 1453; DB 14; Length 381;
Best Local Similarity 49.6%; Pred. No. 6,46e-251;
Matches 193; Conservative 89; Mismatches 96; Indels 11; Gaps 9;

Db 1 mgmraaltgfav1ml1qscayk1vcytswsgyregvgsflpda1qpfcthl1ysfan 60
QY 1 MVRSVAMAGFVLMIPWGSAAKLVCFETWMAQYRQGEARFLPKDLPISCTHLIYAFAG 60
Db 61 tssdm1stwmndesnydk1kn1ktrn1k1lsvggwk1gkrfse1asntert1af 120
QY 61 MT-NHQLSTTWMNDTLQFENGKMKPKLKTLLAIGWNESTQKFTDMVATANNROTFF 119
Db 121 vrvapfl1rsgyfdg1d1awlyp-r-1r-dkyfst1k1elnaeftevg-pgrek1l 174
QY 120 VNSAIRFLKRXSPDLDDMEYPSGSGSPAVDKERFTTLVODLANAFQEOAQTSGKERLL 179
Db 175 lsaalsagkvaldrgyda1sqh1d1f1n1mtydfhgwrgtqghsp1ffrqgktdrrf 234
QY 180 LSAAPVAGQTVYDAEYVDKIAQNLDVFNLAAYDFHGSWEKVTGHNSPLYKROESGAAS 239
Db 235 ysnvnyavgyml1lgaask1lmg1p1tfgk1stlaas-engl1ap1sg1p1rftk1eag 293
QY 240 SLNWDAAVQWMLQKGTSPASKTLIGMPYGRSFTLLASSSDTRVGAPATGSGPTKEGGM 299
Db 294 t1ayye1cd1k1aevh1r1snek1v1fat1gngw1gyeh1esv1n1k1g1f1k1ek1lagamw 353
QY 300 MLAYYEVCSW-KGATKQRIQODQKVPYIRRDNONWGFDDVESFKTKVSTLKQKGLGAGAWMA 358
Db 354 aldllddfgq-tcqpkef1plnaikdala 381
QY 359 ALDLDDPAGFSCNQG-RYPLIQTLROELS 386

RESULT 4
ENTRY I48271 #type complete
TITLE BRP39 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I48271
REFERENCE I48271
#authors Morrison, B.W.; Leder, P.
#journal Oncogene (1994) 9:3417-3426
#title neu and ras initiate murine mammary tumors that share genetic markers generally absent in c-myc and int-2-initiated tumors.
#cross-references MVID:95060797
#accession I48271
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-381 ##label RES
#cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```







Db	147	weyr-v-gv-a-e-eh-aklyeamktafveaktsqgrlllltaavaagxgldgsyrv	200
Qy	139	WIPSSQSPVADKRRFTTLVQDLANAAQOEAQTSIGKERLLLSAAVPAQGRYVAGXEVD	198
Db	201	slgkfdllfmsyvlbhswekndvlbklklnptkgvsgyglftfeaaqywaakgmpe	260
Qy	139	KIAQNLDFVNLMAADFHSMEKRWYGHNSPLTKROEESGAASLVVDAVQOOLQKGPAS	258
Db	261	klllgipmyagwclndpnsctatagaasrpsaaktinpagaasywclkylklegketv	320
Qy	259	KLILGMPYRGNSFTLIASSSDTRVGAAPAGSOTPGPFTEGGMALYEVCSN-KKATGQRI	317
Db	331	hgeygvagymvkgdqwyrygneeetlrklmkwlkexyggafawaldiddftgkscgkyp	380
Qy	318	ODOKV-PYIFPDNQVNGVDVSEFXTKYSYLKOKIGLGAAMWALDLDDFAFGSCNGRYP	376
Db	381	llnalsel 389	
Qy	377	LIQTLROEL 385	
RESULT	10		
ENTRY	A56596	#type complete	
TITLE	chitinase (EC 3.2.1.14) - tobacco hornworm		
ORGANISM	formal_name Manduca sexta #common_name tobacco hornworm		
DATE	11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 03-May-1996		
REFERENCES	A56596		
AUTHORS	Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.		
#journal	Insect Biochem. Mol. Biol. (1993) 23:693-701		
#title	Sequence of a cDNA and expression of the gene encoding		
#cross-references	epidermal and gut chitinases of Manduca sexta.		
#accession	MDID:93357793		
#status	A56596		
#molecule_type	preliminary		
#residues	mRNA		
#cross-references	1-554 #label KPA		
#experimental_source	NCBI:136417, NCBI:136418		
#note	Sequence extracted from NCBI backbone		
KEYWORDS	glycosidase; hydrolase		
SUMMARY	#length 554 #molecular_weight 62203 #checksum 4659		
Query Match	29.8%; Score 1025; DB 12; Length 554;		
Best Local Similarity	38.5%; Pred. No. 5,30e-167;		
Matches 161; Conservative	106; Mismatches 126; Indels 25; Gaps 19;		
Db	17	vgsdstarivcgyfsmwavyrpgvgygdedlpreveckthlysflygtegneavllldpel	76
Qy	16	IPWGSAAKLVCFTTMAAGYRQGEARFLPKDLDPLSLCHTLIIAFAFGMN-H-Q-LST-TW	71
Db	77	dvdkngfnfscslshsbvskfmwavgwaegsskyslmvaqsktrmsflsrsvsflky	136
Qy	72	N-DETLVGEFNGLKKMNPKLKTLLAIGCMNFSSTOKFTDMVATANNROTENVSAIRFRKY	130
Db	137	dldgdldlwevpgaadsgsfkdkflylvgelzrf---lrv-gygwel-teayplan	191
Qy	131	SFDGIDLMEWETPGS--QSSPAVDKERFTTLVQDLANAAQOEAQTSIGKERLLLSAAVPAQ	188
Db	192	flrimegynhvelcegldeahmsydlrgnwagfadvshpilykxphdgdwayeklnvnglh	251
Qy	189	TYVDAGYEVDKIAQNLDFVNLMAADFHSMEKRWYGHNSPLTKROEESGAASLVVDAVQ	248
Db	252	lweekgcpsnklvlglpfygrsftlsagmnnylgftlnkaagsgdpaptytnatqfayy	311
Qy	249	OMLQGTGTPASKRILILCMPTYGRSFTLIASSSDT-RVGA--P-ATGSGTGPPEPKEGCMALY	304
Db	312	elctevvdgdsgskwkegkcpayygygtwvygyedprsvelkmvlykcdygyantam	371
Qy	305	EYCS--WKG-A--TKQRIODQKVPYIFPDNQVNGVDVSEFXTKYSYLKOKIGLGAAMWA	359
Db	372	ldmdfdggl-cge-knplikllhkms--syvpphntentpfbewarpsptpsdpe	426

QY	360	LDLDFAFSCNQGKRYPIQTLROELSLUPY-LPSPGTRPELVKPGCPGSEPEHGPSPOQ	416
RESULT	11		
ENTRY	A38368	#type complete	
TITLE	chitinase (EC 3.2.1.14) precursor	- Bacillus circulans	
ORGANISM	#formal_name	Bacillus circulans	
DATE	28-Jun-1991	#sequence_revision 28-Jun-1991	#text_change
ACCESSIONS	A38368		
REFERENCE	A38368		
authors	Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka, H.		
#journal	J. Biol. Chem. (1990) 265:15659-15665		
#title	Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed its evolutionary relationship to Serratia chitinase and to the type III homology units of fibronectin.		
#cross-references	MUID:90368776		
#accession	A38368		
##status	preliminary		
##molecule-type	DNA		
##residues	1-699	#label WAT	
##cross-references	GB:J05599		
CLASSIFICATION	#superfamily	fibronectin type III repeat homology	
KEYWORDS	glycosidase; hydrolase		
SUMMARY	#length 699	#molecular-weight 73677	#checksum 6757
Query Match	20.3%;	Score 700;	DB 10; Length 699;
Best Local Similarity	40.4%;	Pred. No. 2,24e-104;	
Matches 120;	Conservative 72;	Mismatches 85;	Indels 20; Gaps 18;
Db	144	qlnkikqnpnlkllisvgwtwn-rfsdtaatacrevfansavdflkynfdgyld	202
QY	79	EFNGIKKNPKIKTLAIGGNFSTOFETDWAANNKQFVNSAIRFLKRYSPDGLD	138
Db	203	weypvggldgnsrpedknytllekirekldaagvgdk-kyllltiasgasatya-a	260
QY	139	WEYP-GS-QG-S-PAVDKERFTLVODLANAFQOEKOTSKERELLISAAPVMEQTVDA	193
Db	261	ntelakiaivdwimltydfingawqklsahnaplnydpaaasaagvdpdantfnvaagq	320
QY	194	GVEVDKINQNDPYNLMAVDHSGMEKNTGHSPL-YKRQE-ESGAA-A-SLVNDAVQO	249
Db	321	hldagvpaakivlvvfyigwd-gcsaqgn-9gyqtcgssvsgvteagsfdydlean	378
QY	250	WLQKTPPAKSKILLGMPTRGSRFTLASSSDRVCAPATGSCNPGFTRKEGMLAYEY-CS	308
Db	379	ylnkgyrryndcakvylynaankffisyydeesvgyktaylksxglqgmfwel	435
QY	309	W-K-GATKQRIQDQKPYIFRD-NQW-VGPDVDESFRKTVSKQKGLGAMVWL	360
RESULT	12		
ENTRY	S51591	#type complete	
TITLE	chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) P2 precursor		
ORGANISM	#pathogenesis-related - common tobacco		
DATE	15-Jul-1995	#formal_name Nicotiana tabacum #common_name common tobacco	
ACCESSIONS	S51591	#sequence_revision 08-Sep-1995	#text_change
REFERENCE	S51591		
authors	Helicz, T.; Second, S.; Kaufmann, S.; Geoffroy, P.; Prasad, V.; Brunner, F.; Frilley, B.; Legrand, M.		
#journal	Mol. Gen. Genet. (1994) 245:246-254		
#title	Molecular characterization of a novel tobacco pathogenesis-related (PR) protein: a new plant chitinase/lysozyme.		
#accession	S51591		
##molecule-type	mRNA		
##residues	1-378	#label HEI	
##cross-references	EMBL:X78325		

[illegible]

QY	78	QEFNQLKKNMNRKLTLLAIGWMNSTQKFTDMDVATANNRQTFVNSAIRFLRKSSFSQSLD	137
Db	170	dweyp-addtqatmvlllkeirsqldayaayq-apy-yhflislaapagehy-sflhm	225
QY	138	DMEYSGSGSPAVPKERFTTLVQDLANAFQCEAQTSGKERLLLSAAVPAGOTYVDAGYE	197
Db	226	sdlgavldyvmalmaydyagwswsjsgyndanlfampspnasp-yntqgaikdyikgvpa	284
QY	198	DKIAQNLDFVNLMAVDFEGSGWEKTYGHNSPLYKROESGAASLNVDAVQOVLQKSTPA	257
Db	285	skivjgmpllygraf-esteg-1gqytsq4ys-gsw-englwdy-kvlpkagatcyqd	336
QY	258	SKLLIGMTRYGRSFTLASSSDTRRGAPATGSGTGPPTKEGMLAYEVCSSMKGATQ--	315
Db	337	svagayysdpskskelisfdtrpdmntkvsylknlq1gsgmfw	379
QY	316	RIDQKVPYIFRDNQWGFDDVESFKTKVSYLKQKGLGAMVW	358
RESULT	14		
ENTRY			
TITLE		JQ1975 #type complete	
ORGANISM		chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus	
DATE		(Aphanocladium album)	
		#formal.name Aphanocladium album	
		03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change	
ACCESSIONS		JQ1975	
REFERENCE		JQ1975	
#authors		Balseau, P.L.; Lafay, J.F.	
#journal		Gene (1992) 120:243-248	
#title		Primary structure of a chitinase-encoding gene (chl1) from the filamentous fungus Aphanocladium album: Similarity to bacterial chitinases.	
#accession		JQ1975	
#molecule_type		mRNA	
#residues		1-423 #label BLA	
#cross-references		GB:X64104	
COMMENT		This enzyme is essential for the degradation of insoluble chitin.	
GENETICS			
#gene		chl1	
#introns		47/2: 80/2: 97/1	
CLASSIFICATION		#superfamily Streptomyces chitinase chl40	
KEYWORDS		glycosidase; hydrolase	
FEATURE			
1-34			
35-423		#domain signal sequence #status predicted #label SIG\	
SUMMARY		#length 423 #molecular-weight 46072 #checksum 8802	
Query Match		16.8%; Score 579; DB 5; Length 423;	
Best Local Similarity		36.6%; Pred. No. 1,66e-8;	
Matches 102; Conservative		64; Mismatches 95; Indels 18; Gaps 16;	
Db	115	lkkgrnmkvmlslgyvwt-nppaaassatktktagsvvgmkwgtfdgldiddeyr	173
QY	83	LKKMPKRLKTLAIGGMFSTQKFTDMVATANNRQTFVNSAIRFLRKYSFDDGLDMEYR	142
Db	174	-adatqagmvl1lgavvseldsyaayq-akg-hhflislaapagaynyk1kfa-e-1g	228
QY	143	GSGGSPAYDKERFTTLVQDLANAFQCEAQTSGKERLLLSAAVPAG-QTYDAGYVDKIA	201
Db	229	kvldylnlmaydyagwswnytcghdanlyanpnpnatp-yntddavgaylngvpankiv	287
QY	202	QNLDFVNLMAVDFHSGWEKTYGHNSPLYKROESGAASLNVDAVQOVLQKGPASKLI	261
Db	288	jgmpllygrsfq-qt-eg-ikpynglgs-gsw-englwdykalp-agaatykcdtak	339
QY	262	IGMPYGRSFTLASSSDTRRGAPATGSGTGPPTKEGMLAYEVCSSMKGATKORIODK	321
Db	340	gcysydstkclisfdtpamistvswlkykg1gsgmfw	378
QY	322	VPIYF-RD-NQWVGFDVYESKTYVSYLKQKGLGAMVW	358

RESULT 15

ENTRY S51369 #type complete  
TITLE Chitinase - fungus (Trichoderma harzianum)  
ORGANISM #formal\_name Trichoderma harzianum  
DATE 15-Jul-1995 #sequence\_revision 19-Apr-1996 #text\_change 03-May-1996

ACCESSIONS  
REFERENCE

#authors

Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell, A.; Pinfor-Toro, J.A.

#journal

Curr. Genet. (1994) 27:83-89

#title

Cloning and characterization of a chitinase (CHT42) cDNA from the mycoparasitic fungus Trichoderma harzianum.

#accession

S51369

##status

preliminary

##molecule\_type

DNA

##residues

1-423

##length

423

##molecular\_weight

46056

##checksum

2000

SUMMARY

Query Match 16.48; Score 564; DB 12; Length 423;  
Best Local Similarity 36.08; Pred. No. 1.07e-78;  
Matches 102; Conservative 70; Mismatches 94; Indels 17; Gaps 13;

Db 111 qlfk-vkkanrglkvlslsgwtst-nfpaastdanrknfaktatltfmkdwgfdgldi 168  
QY 78 QEFNGLKKNPKLKTLLAIGGNFSTQKFTDMATANNROTFTVNSAIRFLRKISFDGLD 137

Db 169 dweyp-adatqaasmlllkevrsgdayaay-apy-yhflitlaapagkday-sk1r1 224  
QY 138 DMEYPSGSGSPAVDKERFTLVQDLANAFQEOAOTSKEKRLLSAAYVAGOTYVDAGYEV 197

Db 225 adlqgvldylnmaydyagsfspltyhdanlfnpsnpnatp-ntdsavkdyingvpa 283  
QY 198 DKIAQNLDFVNLMAAYDEFGSMKEKVTGHSNPLYKROESGAASLNVDAVQOOWLQKCTPA 257

Db 284 nklvlgmplygrsf--qntag--lqgtyngvgs-gsw--eaqlwdykaklpk-agatvgyd 335  
QY 258 SKLLGMPYTGSRFTLASSDTRGAPATGSGTRGPFTKEGMLATYEVCSSWKGATKORI 317

Db 336 svakgyysynsatkelisfdtpdmintkvaylkalqigsfmfw 378  
QY 318 ODQKVPIFRD--NQWVGFDVESFKTKVSYLKGKLGAMVW 358

Search completed: Fri Jun 27 16:57:26 1997  
Job time : 72 secs.



Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

```

Title: >US-08-663-018A-4
Description: (1-466) from US08663618A.pep
Perfect Score: 3440
Sequence: 1 MYRSVAMAGMVLMLTPWS.....QDSCTGLVPSNSCKCTWN 466

```

Searched: 59021 seqs, 21210388 residues

Database:

1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 50.039; Variance 84.025; scale 0.596

**Pred. NO.** Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result	Score	Query Match	Length	DB ID	Description	Pred. No
No.						

1	1513	44.0	383	4	GP39_HUMAN	CARTILAGE GLYCOPROTEI	0.00e+00
2	1080	31.4	504	2	CHIT_BROMA	ENDOCHITININASE PRECURS	4.93e-21
3	1055	29.8	554	2	CHIT_BROMA	ENDOCHITININASE PRECURS	1.84e-21
4	700	20.3	699	2	CH11_BACCI	CHITININASE A1 PRECURSOR	3.66e-13
5	579	16.8	423	2	CH11_ALPHA	CHITININASE I PRECURSOR	3.08e-10
6	554	16.4	423	2	CH14_TRIHA	442 KD ENDOCHITININASE	1.62e-99
7	548	15.9	427	2	CH11_COCCI	ENDOCHITININASE I PRECU	6.10e-96
8	440	12.8	563	2	CH1A_SERNA	CHITININASE A PRECURSOR	1.04e-70
9	398	11.6	551	2	CHIT_MPAC	PROBABLE ENDOCHITININS	4.28e-61
10	348	10.1	820	2	CH1A_ALTSO	CHITININASE A PRECURSOR	7.81e-50
11	333	9.7	499	2	CH1B_SERNA	CHITININASE B PRECURSOR	1.66e-46
12	304	8.8	619	2	CH1T_STRLI	CHITININASE C PRECURSOR	3.91e-42
13	280	8.4	610	2	CH1T_STRLP	CHITININASE 63 PRECURSO	4.24e-37
14	265	7.7	1146	5	KTYA_KLUDA	KILLER TOXIN ALPHA AN	5.77e-32
15	172	5.0	36	4	GP39_SOVIN	CARTILAGE GLYCOPROTEI	4.73e-13
16	172	5.0	83	3	EDOS_PANAL	ESTRADIOL-DEPENDENT O	4.73e-13
17	154	4.5	541	11	YKCS_CAEEL	HYPOTHETICAL_55.6 KD	9.26e-10
18	117	3.4	597	2	CH1X_STROI	EXOCHITININASE I PRECUR	1.46e-03
19	110	3.2	497	2	CH1D_BACCI	CHITININASE D PRECURSOR	1.66e-02
20	102	3.0	452	4	YFK2_RHOSH	COPROPORPHYRINOGEN II	2.35e-01
21	99	2.9	289	11	YFK2_YEAST	HYPOTHETICAL_31.9 KD	6.04e-01
22	100	2.9	364	11	YFK3_YEAST	HYPOTHETICAL_40.9 KD	4.44e-01

45	PER	2.7	993	11	YAL1_SCHRO	PUTATIVE FAMILY 31 GL	2.81e+00
44	PGC	2.7	344	4	GC02_PICU	GTP CYCLOHYDROLASE II	2.81e+00
43	YEA	2.7	295	11	Y08	HYPOTHECTAL 35.1 KD	3.78e+00
42	YEA	2.7	267	3	EBAG_F1AP	ENDO-BETA-N-ACETYLGLU	3.78e+00
41	YEA	2.7	99	10	Y150_NPVAC	HYPOTHECTAL 11.2 KD	3.78e+00
40	YEA	2.8	1409	2	C0P1_DROME	COP1 PROTEIN.	2.08e+00
39	YEA	2.8	1333	10	VGR1_MOUSE	VASCULAR ENDOTHELIAL	2.08e+00
38	YEA	2.8	1045	7	PRIS_SERNA	EXTRACELLULAR SERINE	2.08e+00
37	YEA	2.8	563	5	K3CF_HUMAN	KERATIN, TYPE II CYTO	8.30e-01
36	YEA	2.8	563	5	K3CE_HUMAN	KERATIN, TYPE II CYTO	2.08e+00
35	YEA	2.8	563	5	K2CC_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
34	YEA	2.8	563	5	K3CA_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
33	YEA	2.8	563	5	K3CB_HUMAN	KERATIN, TYPE II CYTO	1.13e+00
32	YEA	2.8	554	5	HYES_MOUSE	SOLUBLE EPOXIDE HYDRO	1.54e+00
31	YEA	2.8	554	5	HYES_RAT	SOLUBLE EPOXIDE HYDRO	1.54e+00
30	YEA	2.8	492	9	SYK_THERT	LYS17-TRNA SYNTHETASE	8.30e-01
29	YEA	2.8	422	10	IMDC_ECOLI	UMC PROTEIN.	2.08e+00
28	YEA	2.8	353	11	Y0HT_BACSI	PUTATIVE PEPTIDASE IN	2.08e+00
27	YEA	2.8	297	5	HXU1_HUMAN	HOMEOBOX PROTEIN MSX-	1.54e+00
26	YEA	2.9	1070	1	ALPH1_CANTN	ALPHA-GLUCOSIDASE PHE	3.23e-01
25	YEA	2.9	1045	7	PRIT_SERNA	EXTRACELLULAR SERINE	3.23e-01
24	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
23	YEA	2.9	398	9	SPER_STRY	EXOXYN TYPE B PRECU	4.44e-01
22	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
21	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
20	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
19	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
18	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
17	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
16	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
15	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
14	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
13	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
12	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
11	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
10	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
9	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
8	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
7	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
6	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
5	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
4	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
3	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
2	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01
1	YEA	2.9	958	1	AMWG_SCHOC	EXOXYN TYPE B PRECU	4.44e-01

RESULT	1
ID	GP39_HUMAN
	STANDARD;
	PRT;
	383 AA

DT	01-JUL-1993	(REL. 26, CREATED)
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)
DE	CARTILAGE GLYCOPROTEIN-39	PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).
OS	HOMO SAPIENS (HUMAN).	
OC	EDUAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	
CC	EDUAROTIA; PRIMATES.	
RN	[1]	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RC	TISSUE: CARTILAGE.	
RX	MEDLINE; 94064658.	
RA	HAKALA B.E., WHITE C., RECKLIES A.D.;	
RL	J. BIOL. CHEM. 268:25803-25810(1993).	
RN	[2]	
RP	SEQUENCE OF 22-45.	
RX	MEDLINE; 90328983.	
RA	NIKKROS P., GOLDS E.E.;	
RL	BIOCHEM. J. 269:265-268(1990).	
CC	-1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO	
CC	RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.	
CC	-1- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL	
CC	CELLS AS WELL AS IN LIVER, UNDETECTABLE IN MUSCLE TISSUES, LONG,	
CC	PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.	
CC	-1- PTM: GLYCOSYLATED.	
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR.	
CC	-1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.	
DR	EMBL; M80927; G348912; ..	
DR	PIR; S10677; S10677.	
DR	PIR; A33162; A33162.	
DR	PROSITE; PS01095; CHITINASE_18.	
KW	GLYCOPROTEIN; SIGNAL.	
FT	SIGNAL	1 21
FT	CHAIN	22 383
FT	CARBOHYD	60 60
FT	POTENTIAL.	
SQ	SEQUENCE	383 AA; 42613 MW; D0A5280F CRC32;
D0	Query Match	44.0%; Score 1513; DB 4; Length 383;
D0	Best Local Similarity	52.5%; Pred. No. 0.00e+00;
D0	Matches	200; Conservative 89; Mismatches 83; Indels 9; Gaps 5;
D0	8	tgfvvlllccscaykivcytwsqyregdscfdalrficthlysfanisdhd 67
D0	8	AGFVLLIMPMGSAKLVCTFTNNAQVROGEARFLPDDLPSCITLIVAFAGNTNQLS 67

[illegible]

RESULT	2	STANDARD;	PRT;	504 AA.
ID	CHIT. BRUMA			
AC	P29030;			
DT	01-DEC-1992 (REL. 24, CREATED)			
DT	01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	ENOCCHITINASE PRECURSOR (EC 3.2.1.14) (MEL ANTIGEN).			
OS	BRUGIA MALAYI.			
OC	EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; SPIRIDIIDA.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE; 92119220.			
RA	FURHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;			
RL	PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).			
CC	-1- FUNCTION: THE MEL ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY			
CC	FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-			
CC	FLARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND			
CC	TRANSMISSION.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF			
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-1- PTM: O-GLYCOSYLATED.			
CC	-1- KNOWN TO BIND CALCIUM.			
CC	-1- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MEL ANTIGEN CORRESPOND			
CC	WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.			
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES).			
DR	EMBL; M73689; G156064; -.			
DR	PIR; A38221; A38221.			
KR	PROSITE; PS01095; CHITINASE_18.			
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;			
KM	REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.			
FT	SIGNAL	1	22	
FT	CHAIN	23	504	ENOCCHITINASE.
FT	DOMAIN	23	400	CATALYTIC.
FT	DOMAIN	401	450	SER/THR-RICH (LINKER).
FT	DOMAIN	407	448	3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT	ACT_SITE	148	148	PROTON DONOR (BY SIMILARITY).
QO	SEQUENCE	504 AA;	55971 MW;	4DAYE5B CRC32;

Query Match	31.48;	Score 1080;	DB 2;	Length 504;
Best Local Similarity	42.58;	Pred. No. 4.93e-225;		
Matches 157;	Conservative	95;	Mismatches 102;	Indels 15; Gaps 11;

26 CYTNNAAQYRQGEARFLPKDLDBSLCTHLIYAFAGMTNHQSLSTT-EWND-E-T----LYQ 78

Db	87	avtklrenppljklvlysgyvnfgfsaifbfjaksackterffiknsaiafirfknmfdgld	146
Qy	79	EFNGLKKNRPfLKLfLLAIGCNEFStOKFTDMVAFANRRQFFVNSAIFLKLKYSFSDGLD	138
Db	147	weyr-v-gv-a-e-h-aklyeamktafveeaktsqkrlllllaavaagqgtdgsyve	200
Qy	139	WEYEGSOSPvYDDErERTTfVQDfLANFQEOAQrSGERfLLSAfVAGQrTYDAGfEYD	198
Db	201	slgknfalfimsydlhgswekvnldhgklhprtkgvsygfifnftfaadyaaskgmpe	260
Qy	199	KfAQNfDfVNLMAfDfPFGSWKfVtGHNfSPLYKRGEESGAASfNLVDAfVOQMfLKGfPAS	258
Db	261	klllgfipmyagwclndpsetaiaaasrpsaaskltpagqtaasweickylkegaketv	320
Qy	259	KfLlGMfPfyGSfSfLLASSfSDfRfVAPfAPfTGSfGPfPFfKEGfMfLAfYfVfVCSfM-KGATfQRI	317
Db	321	hgqegayamvsgdwygdyneetfrrlkmkwllkekygsaafwaiddafcgksckgkppp	380
Qy	318	ODQKv-PIfFDfNQfMVGfDDvESfKfTfVfSLfKQfGfLGCfGAMfVfMLfDDfDfFfGfSCfNGfRfP	376
Db	381	llnaatseel 389	
Qy	377	LIQfTLRQEL 385	

ID	RESULT	3	STANDARD;	PRT;	554 AA.
AC	CHIT1.MANSE				
AD	P36362.				
DT	01-JUN-1994	(REL. 29, CREATED)			
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995	(REL. 31, LAST ANNOTATION UPDATE)			
DE	ENOCCHITINASE	PRECURSOR (EC 3.2.1.14).			
OS	MANDUCA sexta	(TOBACCO HAWKMOTH).			
OC	EMMAROTA	METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.			
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE;	93357793.			
RA	KRAMER K. J., CORPUZ L., CHOI H. K., MUTHUKRISHNAN S.;				
RL	INSECT BIOCHEM. MOL. BIOL.	23:651-701(1993).			
CC	-1- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING PROCESS.				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF				
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.				
CC	-1- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,				
CC	BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH				
CC	INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A				
CC	RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER				
CC	LEVELS SEEN ON DAYS 0, 7 AND 8.				
CC	-1- TISSUE SPECIFICITY: EPIDERMIS AND GUT.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
CC	HYDROLASES).				
DR	EMBL;	U02270; G406049; -.			
DR	EMBL;	S64577; -; NOT_ANNOTATED.CDS.			
DR	HSSP;	P07254; ICTN.			
DR	PROSITE;	PS01095; CHITINASE.18.			
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.				
FT	SIGNAL	1	19	POTENTIAL.	
FT	CHAIN	20	554	ENOCCHITINASE.	
FT	DOMAIN	366	453	SER/THR-RICH.	
FT	ACT_SITE	146	146	PROTON DONOR (BY SIMILARITY).	
FT	CARBOHYD	85	85	POTENTIAL.	
FT	CARBOHYD	303	303	POTENTIAL.	
FT	CARBOHYD	407	407	POTENTIAL.	
FT	CARBOHYD	545	545	POTENTIAL.	
SO	SEQUENCE	554 AA;	62203 MM;	FA8/F8AD CRC32;	

Query Match	29.8%;	Score 1025;	DB 2;	Length 554;
Best Local Similarity	38.5%;	Pred. No. 1,84e-211;		
Matches	161;	Conservative	106;	Mismatches 126;
				Indels 25;
				Gaps 19;

Db 17 vgsdsrarlvcvsnawyrpbgrygiedipvekcthllysfllgtvegsevlldpel 76  
.:|::|||:||| |: :| |::||:| |:|:::|

Matches	120;	Conservative	72;	Mismatches	85;	Indels	20;	Gaps	1
Db	144	qlnkikgtupnlkclisvgywtwn-ifsdvaatacrevfansavdflrkynfdyld	202						
Qy	79	EFNGIKKNNPKLTLAIGGWNFSFOKFTDVAANNQTFVNSAIRFLRKYSFSDGLD	138						
Db	203	weypvsgglDgnskrpedkqnytl  sklraklaaagavdgk-yyll  laasgaty-a	260						
Qy	139	WEYP-GS-QG-S-PADKKEFFTLVODLAAFOEOETSSEKLELLSAAAPACQTYDA	193						
Db	261	ntelaklaivdwimlmetdingawqksahnaplmydpaaasagvdpdanfnvaagag	320						
Qy	194	GVEYDKLNQNDLPYNLMAVDFHSGMEKVTGINSPL-YKROB-ESGA-A-SLNDAAVQ	249						
Db	321	hlIdagvpaaklvjvpyfyrgwd-gcaagagn-ggyqtctgssvgtweagsfdfydean	378						
Qy	250	WLOKQTPASKLILCMPTGRSFTLASSSDTRVGAATGSGPGPFTEKGGMLAYEV-C	308						
Db	379	ylnkngtryyndctakvpylynaankrfisydddeesvgytkaylksqylgamtvel	435						
Qy	309	W--K-GATKQRIQDQKVPYIRD-NQW-VGDDVESEFKTYSYLKQGLGGAMWAL	360						
RESULT	5	STANDARD;	PRT;	423	AA.				
AC	P32470:	CH11 ALPHA							
DT	01-OCT-1993	(REL. 27, CREATED)							
DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)							
DT	01-FEB-1996	(REL. 33, LAST ANNOTATION UPDATE)							
DE	CHITINASE 1	PRECURSOR (EC 3.2.1.14).							
GN	CHT1.								
OS	APHANOCLODIUM ALBUM.								
OC	EDUAROTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).								
RN	[1]								
RC	SEQUENCE FROM N.A.								
RP	STRAIN-ETHM 483;								
FX	MEDLINE; P3013040.								
RA	BLAISEAU P.-L., LAFAY J.-F.;								
RL	GENE 120:243-248(1992).								
RN	[2]								
RP	SEQUENCE OF 35-57.								
FX	MEDLINE; 92136437.								
RA	BLAISEAU P.-L., KUNZ C., GRISON R., BERTHEAU Y., BRIGOO Y.;								
RL	COUR. GENET. 21:61-66(1992).								
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF								
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.								
CC	-1- SUBCELLULAR LOCATION: SECRETED.								
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL								
CC	HYDROLASES).								
DR	EMBL; X64104; G429026; -.								
DR	PIR; J01975; J01975.								
DR	HSSP; P07254; 1CTN.								
DR	PROSITE; PS01095; CHITINASE 18.								
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.								
FT	SIGNAL	1	22						
FT	PROPEP	23	34						
FT	CHAIN	35	423						
FT	ACR SITE	171	171						
FT	SEQUENCE	423	AA;	46058	MM;	88456E6	CRC32;		
Query Match		16.8%;	Score 579;	DB 2;	Length 423;				
Best Local Similarity		36.6%;	Pred. No. 3.08e-103;						
Matches 102;	Conservative	64;	Mismatches 95;	Indels 18;	Gaps 16;				
Db	115	lkkqrmkmvmlsigvwtw-ntpaasasatktftagsavgmekdvgffgdlldweyp	173						
Qy	83	LKKMPPKLTLLAIGGWNFSFOKFTDVAIVNNQTFVNSAIRFLRKYSFSDGLDWEYP	142						
Db	174	-adatqagmvl  qavseidvaayaq-akg-nhflislaapgdnylnklfa-e-lg	228						
Qy	143	GSQGSFPAVDKERTFTIVODLAAHQEOEQTSGERLLLSAIVPAG-QTYVADAGEVDKIA	201						
Db	229	kvlidylimaydygswenylcghdanlyanqpnpnatp-ynctdavgaylngvpanklv	287						





OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC ENTEROBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BARK J.D., LEE S.Y.,  
RA CHO M.J.;  
RL SUBMITTED (JAN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 990 / QMB1466;  
RA JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;  
RL EMBL J. 5:467-473(1986).  
RN [3]  
RP REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RA MEDLINE: 95219379.  
RA PERAKIS A., TEWS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,  
RA VORIAS C.E.;  
RL STRUCTURE 2:1169-1180(1994).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL: L01455; G152818; -.  
DR EMBL: X03657; G46831; -.  
DR PIR: A25090; A25090.  
DR HSSP: P07254; 1CTN.  
DR PROSITE: PS01095; CHITINASE\_18.  
RW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.  
FT SIGNAL 1 23  
FT CHAIN 24 563 CHITINASE A.  
FT DOMAIN 150 563 CATALYTIC.  
FT ACT\_SITE 315 563 PROTON DONOR (PROBABLE).  
FT ACT\_SITE 391 563 PROBABLE.  
FT CONFLICT 76 77 TA -> GP (IN REF. 2).  
FT CONFLICT 139 139 A -> P (IN REF. 2).  
FT CONFLICT 226 226 V -> I (IN REF. 2).  
FT CONFLICT 395 395 P -> A (IN REF. 1 AND 2).  
FT CONFLICT 410 430 PAKKPDATITVNGVALLAQ -> RPSRRRLHGHRRQC  
FT CONFLICT 437 437 AAGQ (IN REF. 1 AND 2).  
FT CONFLICT 464 467 V -> I (IN REF. 2).  
FT CONFLICT 473 473 ATGP -> HRA (IN REF. 2).  
FT CONFLICT 484 484 K -> E (IN REF. 2).  
FT SEQUENCE 563 AA; 60979 MW; 7764DB57 CRC32;  
Query Match 12.8%; Score 440; DB 2; Length 563;  
Best Local Similarity 29.1%; Pred. No. 1,04e-70;  
Matches 94; Conservative 83; Mismatches 123; Indels 23; Gaps 19;  
Db 231 pfaalqkqkytawddpykgnfgqlmaakqahpdlklpslsgwtlsdpff-ngdkvk 289  
QY 57 AFAGMTNHQSLSTENNDLTYO-EPYKGNFGQLMAAKLANPHKLKLPISLGSWTLSDP-fyfmhdv 277  
Db 290 -drftvsgvkeftgkwkffdgvdldwefpgkganpnlgspddgelyvllmlelramldq 348  
QY 115 NKQTVNNAIRLKRKYSF-DGLDLDWEIPESQGS-PAV--DKEPFTLVODLANAFQ 168  
Db 349 lsaetgr-kyeltsaisagkdkidkavynvag--nsmdhiflmsydfygfqlklnlghqt 405  
QY 169 EAQTSGRKRLLSAIVPAQGYVD-AGYEVDKIAQNLDFVNLMAIDFHSME-KYTGHS 226  
Db 406 alnapawpdtayt-tvng-vnallagvkvpgkvvvvgvtmgyrgvtyngynipfcgt 463  
QY 227 PLYKROESGAASLNVDAVQOWLOKGPASKILIGMPTGRSFTLASSDTRVGAAPAT 286  
Db 464 atg-pvkgvkwngivdyrgiaqfmgswgylydataaepvyfkpctgdlifddarsq 522  
QY 287 GSGTGPFTKEGMLAYIVCS-WKATKQ-RIQD-QKVPYIFR-DN-QWGFDDVESK 341  
Db 523 akgyvldkqlyglisfaweldach 545  
QY 342 TKVSYLKQKGLGAGAMWALDLD 364

RESULT 9  
ID CHIT\_NPVAC STANDARD; PRT; 551 AA.  
AC P41684;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).  
OS AUTOGRAPH CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMPV).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C6;  
RA MEDLINE: 94303173.  
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;  
RL Virology 202:586-605(1994).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL: L22858; G559195; -.  
DR PROSITE: PS00014; ER-TARGET.  
DR PROSITE: PS01095; CHITINASE\_18.  
RW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.  
FT SIGNAL 1 17  
FT CHAIN 18 551 PROBABLE ENDOCHITINASE.  
FT ACT\_SITE 305 505 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 173 173 POTENTIAL.  
FT CARBOHYD 444 444 POTENTIAL.  
FT SITE 548 551 PREVENT SEQUESTRATION FROM ER (POTENTIAL).  
SQ SEQUENCE 551 AA; 61368 MW; BDB124DB CRC32;  
Query Match 11.6%; Score 398; DB 2; Length 551;  
Best Local Similarity 29.5%; Pred. No. 4.28e-61;  
Matches 96; Conservative 79; Mismatches 122; Indels 28; Gaps 27;  
Db 221 pwaavqkbpqkysawn-ep-ykgnfgqlmaaklanphklklpislgwtlsdp-fyfmhdv 277  
QY 57 AFAGMTNHQSLSTENNDLTYO-EPYKGNFGQLMAAKLANPHKLKLPISLGSWTLSDP-fyfmhdv 277  
Db 278 -ekrlnfvdsykeftgkwkffdgvdldwefpgkganpnlgspddgelyvllmlelram 336  
QY 113 ANNKOTFNNAIRLKRKYSF-DGLDLDWEIPESQGS-PAV-D-K-E-R-FTLVODLANA 165  
Db 337 lddlaeqt-grv-yeltsaisagykliavvnya-e-agslsgkflmsydfkgaasndcl 392  
QY 166 FQO-BAQTSGRKRLLSAIVPAQ-QTYVDAGYEVDKIAQNLDFVNLMAIDFHSMEKYT- 222  
Db 333 gyqtvtva-ps-wmseelycthyavdaalkqgvdpnkliivgamyrgvtyngyndy 450  
QY 223 GHSNPLYKROESGAASLNVDAVQOWLOKGPASKILIGMPTGRSFTLASSDTRVG 282  
Db 451 fsgtng-psgwtgvdvdyrgldkldnnyyrtldsaagsyvdtkgdlifdsyds 509  
QY 283 APATSGTGPFTKEGMLAYIVCSWKATKQRIQD-QKVPYIFR-KDN-QWGFDDVES 339  
Db 510 vlgkvkyvdrnklyglisfaweldach 534  
QY 340 FTKVSYLKQKGLGAGAMWALDLD 364  
RESULT 10  
ID CHIA\_ALTSO STANDARD; PRT; 820 AA.  
AC P32823;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).  
GN CHIA.  
OS ALTEROMONAS SP. (STRAIN O-7).  
OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC VIBRONACEAE.  
RN [1]



Db 425 gsdgqk-ldaa-dygeaskyldwvnmtydffgawak-ngtaphaspilraygqigqfn 481  
Qy 184 VFAQGYVADAGEVDKIAQNLDVFNLMAYDFHSGMEKVTGHNSPLYKROEESGAHA-SLN 242  
Db 482 tadamakfkxspadk11lgigfygwt-gvtgsap-gglatgpat-gty--eagled 536  
Qy 243 VDAAVQOMLQKGTSPASKILGMPYGRSFTLASSSDTRVGARATGSGTRGPFTKESGMMA 302  
Db 537 y-kvlnscapctgtlag--tayahcgsmwsydtcpatlkskmdwaegq1ggafwefsg 593  
Qy 303 YEEVCSMGKATKORIQDQKVPYIFRDQWVGFDVESFKTKVSYLKQKGLGAMVWALDL 362  
Db 594 dtang 598  
Qy 363 DDFAG 367

RESULT 13  
ID CHIT\_STRPL STANDARD; PRT: 610 AA.  
AC P11220.  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE 63 PRECURSOR (EC 3.2.1.14).  
CN CHTA  
OS STREPTOMYCES PLICATUS.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92192480.  
RA ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;  
RL GENE 111:69-76(1992).  
RN [2]  
RP SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.  
RX MEDLINE: 88087127.  
RA ROBBINS P.W., ALBRIGHT C., BENFIELD B.;  
RL J. BIOL. CHEM. 263:443-447(1988)  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC -1- HYDROLASES).  
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLOLOSE-BINDING DOMAIN  
CC (CBD).  
DR EMBL: M82804; G153216; -;  
DR EMBL: M18397; G153209; -;  
DR PIR: A29912; A29912.  
DR PIR: JH0573; JH0573.  
DR HSSP: P07254; ICTN.  
DR PROSITE: PS00018; EF\_HAND.  
DR PROSITE: PS00561; CBD\_BACTERIAL.  
DR PROSITE: PS01095; CHITINASE\_18.  
KV HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.  
FT SIGNAL 1 30  
FT CHAIN 31 610 CHITINASE 63.  
FT DOMAIN 35 140 CELLOLOSE-BINDING.  
FT DOMAIN 148 229 FIBRONECTIN TYPE-III.  
FT DOMAIN 236 610 CATALYTIC.  
FT ACT\_SITE 383 383 PROTON DONOR (BY SIMILARITY).  
FT CONFLICT 3 3 F->I (IN REF. 2).  
SQ SEQUENCE 610 AA; 63974 MW; 2F5E8E35 CRC32;

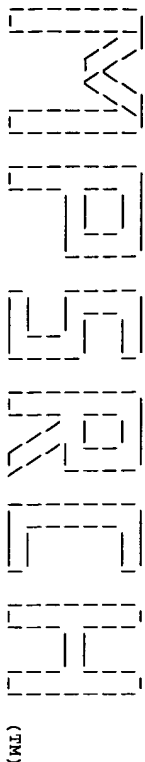
Query Match 8.4%; Score 290; DB 2; Length 610;  
Best Local Similarity 26.7%; Pred. No. 4,24e-37;  
Matches 79; Conservative 62; Mismatches 134; Indels 21; Gaps 16;

Db 320 fnglnlkaeyphikllyfsgwtwsgg-fpdavxnpaafakschldlvedprwadvfqgi 378  
Qy 77 YQEFNLKMKMFKLTLAIGMNFSTQKFTDMVATANNRQGFVNSAIFRLKRY-S-FDGL 135  
Db 379 dldwepnacyglscdetsapnafsmkamreafgqdyilta-a---vtaadsgdgk-id 433

Qy 136 DIDWYPPSSQG-S--PAVDKERFTLVQDLANAFQGEAQTSGSKERLLLSAANPAQTVVD 192  
Db 434 aa-dygeaskyldwvnmtydffgawak-ngtaphaspilnadygipqgftadamakfk 491  
Qy 193 AGEVXDKAQNLDVFNLMAYDFHSGMEKVTGHNSPLYKROEESGAHA-SLVNDAVQOML 251  
Db 432 skvypadk11lgigfygwt-gvtgsap-gglatgpa-gty--eagledy-kvlnsc 545  
Qy 252 QKGTSPASKILGMPYGRSFTLASSSDTRVGAPAGSGTRGPFTKEGGLAYEVCSSMG 311  
Db 546 patgvaq--tayahcgsmwsydtcpatlkskmdwaegq1ggafwefsgdtng 599  
Qy 312 ATKORIQDQKVPYIFRDQWVGFDVESFKTKVSYLKQKGLGAMVWALDDDFAG 367

RESULT 14  
ID KTYA\_KLULA STANDARD; PRT: 1146 AA.  
AC P09805.  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (REF2 PROTEIN)  
DE KUDOCITINASE (EC 3.2.1.14).  
OS KLUYVEROMYCES LACTIS (YEAST).  
OC PLASMOD BERL-1.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CBS 2359/152;  
RA SOR F., FUKUHARA H.;  
RL CURR. GENET. 9:147-155(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84297209.  
RA STARK M.J.R., MIDEHAM A.J., ROMANOS M.A., BOYD A.;  
RL NUCLEIC ACIDS RES. 12:6011-6030(1984).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85037931.  
RA TOKUNAGA M., KAWAMURA A., HISHINUMA F.;  
RL NUCLEIC ACIDS RES. 12:7581-7597(1984).  
RN [4]  
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.  
RX MEDLINE: 87004569.  
RA STARK M.J.R., BOYD A.;  
RL EMBO J. 5:1995-2002(1986).  
RN [5]  
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.  
RX MEDLINE: 90259069.  
RA BRADSHAW H.D. JR.;  
RL NATURE 345:299-299(1990).  
RN [6]  
RP CHITINASE ACTIVITY OF ALPHA-SUBUNIT.  
RX MEDLINE: 91301161.  
RA BUTLER A.R., O'DONNELL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;  
RL EUR. J. BIOCHEM. 199:483-488(1991).  
CC -1- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE, ALONG WITH  
CC THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE  
CC TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE  
CC TOXIN) TO GAIN ENTRY INTO THE CELL.  
CC -1- P1M: REF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-  
CC SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.  
CC -1- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,  
CC BETA AND GAMMA.  
CC -1- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.  
DR EMBL: X07127; G2829; -;  
DR EMBL: X00762; G2844; -;  
DR EMBL: X01095; G2849; -;  
DR PIR: S07915; S07915.  
DR HSSP: P02877; 1HEV.  
DR PROSITE: PS00026; CHITIN\_BINDING.





Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 27 17:03:11 1997; Maspar time 9.60 Seconds  
423.242 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-663-618A-14  
Description: (1-373) from US08663618A.pep  
Perfect Score: 2721  
Sequence: 1 AKLVCFYFTNMAQRYROGEARF.....YPLIQTLRQELSLPYLPSGT 373

Scoring table:  
PAM 150  
Gap 11

Searched: 92623 segs, 10896596 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

## Database:

a-geneseq26  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19

Statistics: Mean 34.970; Variance 153.103; scale 0.228

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1328	48.8	537	14	R73991	Bovine oviduct specif	7.85e-114
2	1303	47.9	668	14	R73993	Hamster oviduct speci	2.08e-111
3	1303	47.9	718	14	R73992	Murine oviduct specif	2.08e-111
4	881	32.4	321	1	P81342	Polypeptide involved	9.90e-71
5	597	21.9	371	13	R70025	Tobacco chitinase enc	9.86e-44
6	597	21.9	377	13	R70029	Tobacco cluster-A pro	9.86e-44
7	581	21.4	424	12	R63441	Trichoderma harzianum	3.16e-42
8	576	21.2	389	10	R56861	Aphanocladium album m	9.31e-42
9	576	21.2	423	6	R33072	Aphanocladium album p	9.31e-42
10	569	20.9	389	6	R33068	Aphanocladium album m	4.23e-41
11	463	17.0	866	19	W02159	Soluble chitinase	3.32e-31
12	389	14.3	561	12	R64823	Chitinase A	2.18e-24
13	372	13.7	561	9	R48981	Bacterial chitinase-A	7.79e-23
14	271	10.0	238	6	R32547	HANPV IF-1 gene prote	9.19e-14
15	219	8.0	1046	16	W02156	Periplasmic chitodept	3.07e-09
16	139	5.1	25	13	R70751	YKL-40 N-terminal pep	1.21e-02
17	139	5.0	25	11	R70745	YKL-40 N-terminal seq	1.21e-02
18	136	5.0	29	11	R60073	Bovine Oviduct Glycop	2.08e-02
19	101	3.7	1070	4	R21521	Alpha galactosidase.	8.46e+00
20	99	3.6	398	18	R95856	S. pyogenes speb gene	1.17e+01

21	99	3.6	428	16	R86905	Murine BMP-9.	1.17e+01
22	99	3.6	428	6	R31372	Murine bone morphogen	1.17e+01
23	99	3.6	553	14	R80445	Murine soluble epoxid	1.17e+01
24	99	3.6	958	1	P81181	Sequence of glucocamy1	1.17e+01
25	99	3.6	958	2	R07575	Glucocamyase encoded	1.17e+01
26	95	3.5	438	9	R47259	Pre-pro-VGR1.	2.22e+01
27	92	3.4	1356	5	R26999	Novel type III RTK en	3.57e+01
28	91	3.3	19	13	R70752	YKL-40 internal pepti	4.18e+01
29	91	3.3	19	12	R70746	YKL-40 internal pepti	4.18e+01
30	90	3.3	256	17	R97235	Mutu putative oncogen	4.88e+01
31	90	3.3	459	17	R88058	Protein encoded by Mu	4.88e+01
32	90	3.3	575	3	P50018	Sequence encoded by t	4.88e+01
33	90	3.3	709	2	P70423	FIXD nitrogen fixatio	4.88e+01
34	90	3.3	709	3	P50017	Sequence encoded by t	4.88e+01
35	87	3.2	369	18	R98125	Human lymphocyte cell	6.65e+01
36	88	3.2	513	9	R47264	Pre-pro BMP.	6.65e+01
37	88	3.2	513	2	R07311	Human Bone Morphogene	6.65e+01
38	88	3.2	513	6	R32904	Human BMP-6 peptide s	6.65e+01
39	88	3.2	513	7	R36735	Human BMP-6	6.65e+01
40	87	3.2	542	16	R90772	Bacillus steatothermo	7.76e+01
41	88	3.2	856	17	R92750	Human EGF receptor su	6.65e+01
42	86	3.2	2332	8	R43257	Human Factor VIII.	9.04e+01
43	87	3.2	2351	3	P50319	Human antihemophilic	7.76e+01
44	86	3.2	2351	3	P50059	Human factor VIII.C.	9.04e+01
45	86	3.2	2351	13	R78223	Human Factor-VIII.C.	9.04e+01

## ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	R73991 standard; Protein: 537 AA.	
AC	R73991;	
DT	10-JAN-1996 (first entry)	
DE	Bovine oviduct specific glycoprotein.	
KW	Bovine oviduct specific glycoprotein; recombinant production;	
KW	BOGP.	
OS	Bos taurus.	
FH	Key	Location/Qualifiers
FT	Peptide	1..18
FT	/label- sig_peptide	
FT	Peptide	19..537
FT	/label- mat_peptide	
PN	307107979-A.	
PD	25-APR-1995.	
PF	15-AUG-1994; 214227.	
PR	19-AUG-1993; JP-227881.	
PA	(KINO-) KINOSER PEPTIDE KENKYUSHO KK.	
DR	WPI: 95-190179/25.	
DR	N-PSDB: Q90442.	
PT	New DNA encoding an oviduct-specific glycoprotein - useful for	
PT	recombinant protein production in high quantities.	
PS	Clam 4; Pages 7-9; 22pp; Japanese.	
CC	Q90442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).	
CC	The cDNA can be used for the commercial recombinant prodn. of	
CC	BOGP in high quantities.	
CC	Sequence 537 AA:	

Query Match	48.8%; Score 1328; DB 14; Length 537;
Best Local Similarity	51.2%; Pred. No. 7.85e-114;
Matches	187; Conservative 74; Mismatches 97; Indels 7; Gaps 6;
DB	20 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 79
QY	2 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 60
DB	80 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 139
QY	61 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 120
DB	140 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 199
QY	121 klvcyftnwafrspasillprldpflcthlvafasmsnqlypkdqekilypefn 180

```
Db      200 rllffisvlsyrdlngswkvvgghnspflfs-l-pgdcbk-sa-yammyvwqlygyppekll 254
QY      181 ONDFFVNMAADDFGGSWEKVTGNHSPLYKKROEESGAASLWDDAAVOOMLOKGTPASKLI 240

Db      255 mglplygrtfhnlkasqnrlragavgpaspbkylckgaflaglayelccfvrrakkrwmdq 314
QY      241 LGMPPLTGGSFLLASSSDPRVCAAPRTGSGTPPFTKEGMALAYIVCSM-KGATQRQLQDO 299

Db      315 yvvpaifkgkwvywdaldaisfykaflfkrehfgagmwvldlddfrgyfscgtgpfplvt 374
QY      300 KVPPIYFRDNQMVGDDVDVESFKTSYLKOKLGAGAMWALDLDDFAFGSCNOGRXPPIQT 359

Db      375 lnnll 379
QY      360 LROEL 364
```

RESULT	2
ID	R73993 standard; Protein; 668 AA.
AC	R73993;
DT	10-JAN-1996 (first entry)
DE	Hamster oviduct specific glycoprotein.
KW	Hamster oviduct specific glycoprotein; recombinant production;
HM	HOOP.
OS	Cricetulus griseus.
FT	Key
FT	Peptide
FT	/label= sig_peptide
FT	Peptide
FT	/label= mat_peptide
FT	U07107979-A.
PD	25-APR-1995.
PF	15-AUG-1994; 214227.
PR	19-AUG-1993; JP-227881.
PR	(KINO.) KINOSEI PEPTIDE KENKUSHO KK.
DR	WPI; 95-190179/25.
DR	N-PSDB: Q90444.
PT	New DNA encoding an oviduct-specific glycoprotein - useful for
PT	recombinant protein production in high quantities.
PS	Claim 4; Pages 16-18; 22pp; Japanese.
CC	Q90444 encodes R73993 hamster oviduct specific glycoprotein (HOOP)
CC	The cDNA can be used for the commercial recombinant prodn. of
CC	HOOP in high quantities.
CC	Sequence 668 AA;

DB	374	linell	379
	:		
OY	359	TUROEL	364
OS	Mus musculus.		
FH	Key	Location/Qualifiers	
FT	Peptide	1..18	
FT	/label- sig-peptide		
FT	peptide	19..718	
FT	/label- mat-peptide		
PN	T07107979-A.		
PD	25-APR-1995.		
PF	15-AUG-1994.	214227	
PR	19-AUG-1993; JP-227881.		
PA	(KINO-) KINOSEI PEPRIDE	KENKYUSHO KK.	
DR	WPI; 95-190179/25.		
DR	N-PESDB; G90443.		
PT	New DNA encoding an oviduct-specific glycoprotein - useful for recombinant protein production in high quantities.		
PS	Claim 4; Pages 11-16; 22pp; Japanese.		
CC	O90443 encodes R73992 murine oviduct specific glycoprotein (MOGP).		
CC	The cDNA can be used for the commercial recombinant prodn. of MOGP in high quantities.		
SQ	Sequence	718 AA;	

RESULT 4  
ID P81342; standard; protein; 321 AA.  
AC P81342;  
DI 19-OCT-1990 (first entry)  
DE Polypeptide involved in protective mechanisms  
KW Immune response; cell growth.  
PN J63032898-A.







Query Match	20.9%	Score 569	DB 6	Length 389
Best Local Similarity	35.8%	Pred. NO. 4.23e-41	Indels 18	Gaps :
Matches 100	Conservative 66	Mismatches 95	Indels 18	Gaps :
DB	81	Lkqgrmkvwsigycwtst-nfpaasaaatrcrkffagsgvfmkdwfgdgidweyp	139	
QY	62	LKKNNPKKTLTLAIGGNFTGQKFTDMVAATANNRQFFVNSAIRFLTKYSFDDGLDMEFP	121	
DB	140	-adettqagmnvlllgavrselasyaayg-akg-hhflisaapagpdnykiklfa-e-1g	194	
QY	122	GSQGSAPADKERRFTTLVQDLNANFQEAQTSQKRELLLSAAYVAG-QTYVDAGYEVDKIA	180	
DB	195	kvlvayinlmaydygswsnytgndaniyaqpnpnatp-ynctddavqaylmgvgpankiv	253	
QY	181	QNLDFVNLMADDFGWSKEVYTGHNSPLYKQDESGAASLNDAAQVQWLQKGTAPSKLI	240	
DB	254	lgmlygsfsfg-qt--eg-1gkvpngigs-gsw--eng1wdyke1kp-agatvcddatx	305	
QY	241	LGMPTGGSFTLASSDTRRGAPATGCTGPTTKEGGMALYEVCWSKGATKQRIQDOK	300	
DB	306	gcysydpstkelisfcdpamstskvswlkqk1gsgmfw	344	
QY	301	VPIYF-RD-NQWGFDDVESPKTKVSYLAKQKGGAGMVM	337	
RESULT	11			
ID	W02159	standard; Protein; 866 AA.		
AC	W02159			
DT	14-JAN-1997	(first entry)		
DE		Soluble chitinase.		
KW		Periplasmic chitinodextrinase; periplasmic Beta-N-acetylglucosaminidase		
RW		Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic; catabolism.		
OS	Vibrio furnissii.			
PN	MO0625424-A1.			
PD	22-AUG-1996.			
PF	13-FEB-1996;	U02332.		
PR	13-FEB-1995;	U5-386727.		
PA	(UVOJ ) UNIV JOHNS HOPKINS.			
PI	Baslier B, Chittlaru E, Keyhani N, Roseman S, Rowe C;			
PT	Yu C			
DR	WPI: 96-393335/39.			
DR	N-PSDB: T36390.			
PT	Chitin biosynthetic enzymes end I, exo I and exo II - are			
PT	periplasmic chitinodextrinase(s), periplasmic beta-glucanidase(s) and			
PT	aryl-beta-N-acetylglucosaminidase(s), respectively			
PS	Example 4: Page 79-82; 101pp; English.			
CC	Periplasmic chitinodextrinase (W02159), periplasmic			
CC	Beta-N-acetylglucosaminidase (W02157) and aryl			
CC	Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin			
CC	oligosaccharides with the structure (GlcNAc)n where n is 2 or			



QY 206 PLVKRQESGAAASLNDAAVQWLOKGTASKLILGMPYGRSFTLASSSDRIVGAPAT 265  
 DB 463 hravkgtw--englvdyrglasqfmsgewgyltdgtaeapvfkpstgdlitddarsvq 520  
 QY 266 GSSTPGFTEEGGMLAYVECS-WKCATKQ-RIDQ-QKVYIFR-DN-QWVGFDVDSFK 320  
 DB 521 akkyvldkqlgglfweladn 543  
 QY 321 TKVSTLKQKGLGGAMVALDLD 343

RESULT 14  
 ID R32547 standard; Protein; 238 AA.  
 AC R32547;  
 DT 17-JUN-1993 (first entry)  
 DE HanPV IE-1 gene protein (partial sequence).  
 KW Heliothis armigera nuclear polyhedrosis virus; Ha; NPV;  
 KM insecticide; immediate early-1.  
 OS Heliothis armigera nuclear polyhedrosis virus.  
 PN MO9303144-A.  
 PD 18-FEB-1993.  
 PF 05-AUG-1992; AC0413.  
 PR 05-AUG-1991; AU-007576.  
 PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 PI Christian PD, Dall DJ, Gordon KHJ, Hanzlik TN, Srisantha A.  
 DR WPI: 93-076501/09.  
 N-PSDB: Q35982.  
 PT Insect virus with reduced capacity to occlude viral particles - used  
 PT for controlling proliferation of insect pests without horizontal  
 PT transmission  
 PS Disclosure: Fig 2: 51pp; English.  
 CC The sequence is that encoded by the Heliothis armigera (Ha)  
 CC nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene  
 CC (partial sequence).  
 SQ Sequence 238 AA;

Query Match 10.0%; Score 271; DB 6; Length 238;  
 Best Local Similarity 34.1%; Pred. No. 9.19e-14;  
 Matches 58; Conservative 43; Mismatches 51; Indels 18; Gaps 14;

DB 44 erdkflsnpgwrefvltwxf-fdgididwefpgskgpnvgdverdnmtyiallgelra 102  
 QY 94 NRQTFVNSAI-R-F-L-RKISFDGLDLDWETPESQGS-PAV-DKER---FTTLVODLAN 143  
 DB 103 midq-vqignturlelttaasgldktaav-ncdraqyldkifmsydfkgawsntldg 160  
 QY 144 AFQGEAQTSGKERLLLSAAVPAGQTYVDAGYEVDKIAQNLDYVNLMAVDFHGSMEKVT-G 202  
 DB 161 hgtaly--gsaawhpnepytanvavdallagrvnp-kklvlgvamygrgwt 207  
 QY 203 HNSPLVKRQESGAAASLNDAAVQWL-OKGTASKLILGMPYGRSFT 251

RESULT 15  
 ID W02156 standard; Protein; 1046 AA.  
 AC W02156;  
 DT 14-JAN-1997 (first entry)  
 DE Periplasmic chitodextrinase.  
 KW Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;  
 KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;  
 KW catabolism.  
 OS Vibrio furnissii.  
 PN W09625424-A1.  
 PD 22-AUG-1996.  
 PF 13-FEB-1996; U02332.  
 PR 13-FEB-1995; US-386727.  
 PA (UXTD ) UNIV JOHNS HOPKINS.  
 PI Baasler B, Chittiaru E, Keyhani N, Roseman S, Rowe C;  
 PI Yu C;  
 DR WPI: 96-393335/39.  
 DR N-PSDB: T36387.  
 PT Chitin biosynthetic enzymes end I, exo I and exo II - are

PT periplasmic chito:dextrinase(s), periplasmic beta-Glucanidase(s) and  
 PT aryl beta-N-acetyl:gluco:amidase(s), respectively  
 PS Claim 2; Page 68-71; 101pp; English.  
 CC Periplasmic chitodextrinase (W02156), periplasmic  
 CC Beta-N-acetylglucosaminidase (W02157) and aryl  
 CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
 CC oligosaccharides with the structure (GlcNAc)n where n is 2 or  
 CC higher, by contacting them with soluble chitin. The enzymes are  
 CC encoded by the genes endI, exoI and exoII respectively. They are  
 CC all genes involved in the catabolic pathway of chitin.  
 SQ Sequence 1046 AA;

Query Match 8.0%; Score 219; DB 19; Length 1046;  
 Best Local Similarity 33.3%; Pred. No. 3.07e-09;  
 Matches 48; Conservative 40; Mismatches 42; Indels 14; Gaps 10;

DB 454 etfadaavemmkkyrfidgididrlisnldgltgnpddtfsesrraylmnsyhelmrvir 513  
 QY 96 QTFVNSAIRLKRKTSFDGLDLD-W--EYPSQSPV-VP-KE-RFTTV---QDLNARQ 146  
 DB 514 ekldvasagdgvyhmltlaaps-ayllrgmetavtgyldvvnlnsydlhgandhvhg 572  
 QY 147 QEAQ-TSGKE--RLLSAAVPAGQTYVDAGYEVDKIAQNLDYVNLMAVDFHGSMEKVTG 203  
 DB 573 naalydtgkdselaqg-nvygtag 595  
 QY 204 NSPLVKRQESGAAASLNDAAVQ 227

Search completed: Fri Jun 27 17:04:25 1997  
 Job time : 74 secs.



Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

```

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm

Run on:      Fri Jun 27 17:01:00 1997;      MasPar time 17.84 Seconds
Tabular output not generated.      596.027 Million cell updates/sec

```

```

Title: >US-08-663-618A-14
Description: (1-373) From US08663618A.pep
Perfect Score: 2721
Sequence: 1 AKLVCTYTNNAQRQROGEAR.....YLITLTQRELSPYLDISGT 373

```

Scoring table: PAM 150  
Gap 11

Searched: 89912 seqs, 28507787 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
Database:
plrs0
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics:      Mean 47.350;  Variance 104.053;  scale 0.455
```

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred.	No.
	1	1488	54.7		383	13	A49562	cartilage glycoprote	8.90e-25	
	2	1448	52.5		383	16	S51327	heparin-binding glyco	2.76e-23	
	3	1417	52.1		381	14	S61550	BRP3 protein - mus	3.51e-23	
	4	1417	52.1		381	14	I48871	BRP3 protein - mous	3.51e-23	
	5	1345	49.4		399	5	S27879	secretory protein ym	2.04e-22	
	6	1378	48.8		537	14	S57197	oviduct-specific gly	3.60e-22	
	7	1338	48.4		539	14	I46870	estrogen dependent o	2.92e-21	
	8	1179	43.3		483	12	A53518	chitinase (EC 3.2.1.	8.78e-19	
	9	1086	39.9		504	12	A38321	chitinase (EC 3.2.1.	3.90e-17	
	10	1015	37.3		554	12	A56596	chitinase (EC 3.2.1.	1.05e-16	
	11	698	25.7		699	10	A38368	chitinase (EC 3.2.1.	2.18e-10	
	12	582	21.8		378	5	S51591	chitinase (EC 3.2.1.	2.86e-82	
	13	582	21.4		424	5	S47133	chitinase (EC 3.2.1.	4.43e-80	
	14	576	21.2		423	5	UQ1575	chitinase (EC 3.2.1.	5.50e-79	
	15	561	20.6		423	12	S51369	chitinase - fungus (	2.95e-76	
	16	546	20.1		427	12	UC4565	chitinase (EC 3.2.1.	1.56e-73	
	17	446	16.4		563	16	S60651	chitinase precursor	1.45e-55	
	18	404	14.8		561	9	A25090	chitinase (EC 3.2.1.	3.83e-48	
	19	345	12.7		799	8	PC4106	chitinase (EC 3.2.1.	6.99e-38	
	20	344	12.6		499	8	S52422	chitinase (EC 3.2.1.	1.04e-37	
	21	344	12.6		820	16	A40633	chitinase (EC 3.2.1.	1.04e-37	

22	337	12.5	499	9	S04856	chitinase (EC 3.2.1.	1.65e-3
23	314	11.5	511	11	S61166	probable membrane pr	1.37e-3
24	287	10.5	610	10	TJ0573	chitinase (EC 3.2.1.	4.76e-2
25	266	9.8	452	12	JC4038	47k glycoprotein - f	1.43e-2
26	263	9.7	1146	11	S07915	Rf2 glycoprotein - yeast	4.45e-2
27	246	9.0	413	5	JC2135	chitinase (EC 3.2.1.	2.62e-2
28	229	8.4	55713	5	S57715	chitinase (EC 3.2.1.	1.40e-1
29	172	6.3	36	14	A27682	39k whey protein - b	7.80e-1
30	172	6.3	83	13	A37954	estradiol-stimulated	7.80e-1
31	136	5.0	29	14	S57204	oviduct-specific gly	8.95e-0
32	136	5.0	29	14	S55693	oviduct-specific gly	8.95e-0
33	121	4.4	597	5	S53848	chitinase (EC 3.2.1.	7.93e-0
34	121	4.4	597	5	S53848	chitinase (EC 3.2.1.	7.93e-0
35	112	4.1	647	15	S37623	homologues with orf	1.02e-0
36	110	4.0	488	10	A41961	chitinase (EC 3.2.1.	1.77e-0
37	104	3.8	305	9	S28440	probable copropoxy	8.88e-0
38	104	3.8	422	4	ZWEC3	umuc protein - Esche	8.88e-0
39	100	3.7	364	11	S45325	hypothetical protein	2.52e+0
40	101	3.7	562	3	KR0UEB	keratin, 56k type II	1.94e+0
41	101	3.7	662	11	S61193	probable membrane pr	1.94e+0
42	101	3.7	1045	2	SUSEM1	serine proteinase (E	1.94e+0
43	101	3.7	1070	1	S19686	alpha-glucosidase (E	3.25e+0
44	99	3.6	398	5	A37768	streptococcal pyro	3.25e+0
45	99	3.6	958	1	JN0102	glucan 1,4-alpha-glu	3.25e+0

## ALIGNMENTS

	ENTRY	1	A49562	#type complete	
	TITLE		cartilage glycoprotein gp39 precursor - human		
	ALTERNATE_NAMES		39k synovial protein		
	ORGANISM		#formal_name Homo sapiens #common_name man		
	DATE		23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 03-May-1996		
	ACCESSIONS		A49562; S10677; A33162		
	REFERENCE		A49562		
	#authors		Hakala, B.E.; White, C.; Recklies, A.D.		
	#journal		J. Biol. Chem. (1993) 268:25803-25810		
	#title		Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family.		
	#accession		A49562		
	#status		Preliminary		
	#molecule_type	mRNA			
	#residues	1-383	#label	HAK	
	#cros-references	GB:M80927			
	REFERENCE	S10677			
	#authors		Nyirkos, P.; Golds, E.E.		
	#journal		Biochem. J. (1990) 265:265-268		
	#title		Human synovial cells secrete a 39 kDa protein similar to a bovine mammary protein expressed during the non-lactating period.		
	#cros-references	MUID:90328983			
	#accession	S10677			
	#molecule_type	protein			
	#residues	22-40,'X',42-45	#label	NW2	
	KEYWORDS		cartilage; extracellular protein; glycoprotein		
	FEATURE	1-21			
		22-383			
	SUMMARY		#length 383	#molecular-weight 42613	#checksum 2942
D	Query Match		54.7%	Score 1488;	DB 13; Length 383;
B	Best Local Similarity		53.3%;	Pred. No. 8,906-251;	
O	Matches	195;	Conservative	83;	Mismatches 79; Indels 9; Gaps 5;
D	23	klyvyysvqyrgsgscfpdalrftclcthiyfsafnsndhidtwendvtlygnlnt	82		
		:           :			
O	2	KLVCFYFNMAQVHQSGARFLPKDLDPSLCITHLIVFAAGMTWHQLSTEWNDETLLYOEPNG	61		
D	83	lknmpnlklilsvgswnfigsqrfshiaantsrttflkswprflrthqfdqldlawlyp	142		

```

OY 62 LKKNPKLTKLTAIGNNEGTOKFTDMVATANNRQTFVNSAIRFLRKSPDGLDMWEXP 121
DB 143 grr-----dkqhtltikemkaefikeaq-pgkqlllssalsagvltlssydlaktsq 196
OY 122 GSGGSAVNDKERFTTLVODLANAFQOEPAQTSKGERLLLSAAVPAQGTYYDAGYEVDK 181
DB 197 hldfslmtydfhgwargtqthsp1frggedasprfndtyavym1rlgapaaklv 256
OY 182 NDFVNLMAVDFHGSWEKVTGHNSPLYKKRDESGAASLVNDAVQOMLOKGTIPASKLIL 241
DB 257 glptfgrsfllaas-etyvqapispgipgrfkeaglayelcdflgavtvlr1gq 315
OY 242 GMPTGRSFTLASSSDTRVGAAPATGSGTGPFTKEGMLAYEVCW-KGATKORIQDCK 300
DB 316 vpyatgngvgyddgsvskvqylkdrqlaagmwaldlddfgscfgqdlr1fplna 375
OY 301 VPIYFDNOMVGFDDVESFRTKVSYLKQKGLGAMWALDLDDEFAGFSNQG-RYPLIOT 359
DB 376 lkda1a 381
OY 360 LRQELS 365

RESULT 2
ENTRY 2 S51327 #type complete
TITLE heparin-binding glycoprotein 38k - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 07-May-1995 #sequence_revision 07-May-1995 #text_change
ACCESSIONS S51327
REFERENCE S51327
#authors Shackleton, L.M.; Mann, D.M.; Millis, A.J.T.
#submission Submitted to the EMBL Data Library, January 1995
#description Identification of a 38kDa heparin-binding glycoprotein
(gp38k) in differentiating vascular smooth muscle cells as
a member of a group of proteins associated with tissue
remodeling.
#accession S51327
#status Preliminary
#residues 1-383 #label SHA
#cross-references EMBL:247803
SUMMARY #length 383 #molecular_weight 42443 #checksum 2907

Query Match 52.5%; Score 1428; DB 16; Length 383;
Best Local Similarity 52.2%; Pred. No. 2.76e-237;
Matches 191; Conservative 81; Mismatches 85; Indels 9; Gaps 5;

DB 23 klvcytswsgyregdgsfda1dp1fclthllysfanssmdlstwmdvlydtlnt 82
OY 2 klvcyftnmaoYRQGEARFLPKDLPSLCTHLIYAFAGMTNHQSTENDETLYOEFN 61
DB 83 lkrrnplkllsvgvnfgsfgrfkslaasntgsrrtfkxvppflrtfthgfdglawisp 142
OY 62 LKKNPKLTKLTAIGNNEGTOKFTDMVATANNRQTFVNSAIRFLRKSPDGLDMWEXP 121
DB 143 grr-----dkrhtltikemkaefivreal-pgterlllsgvasagvaidrgydlasq 196
OY 122 GSGGSAVNDKERFTTLVODLANAFQOEPAQTSKGERLLLSAAVPAQGTYYDAGYEVDK 181
DB 197 hldfslmtydfhgwargtqthsp1frggedasprfndtyavym1rlgapaaklv 256
OY 182 NDFVNLMAVDFHGSWEKVTGHNSPLYKKRDESGAASLVNDAVQOMLOKGTIPASKLIL 241
DB 257 glptfgrsfllaas-tdvapaspgipgrfkeaglayelcdflgavtvlr1gq 315
OY 242 GMPTGRSFTLASSSDTRVGAAPATGSGTGPFTKEGMLAYEVCW-KGATKORIQDCK 300
DB 316 vpyatgngvgyddgsvskvqylkdrqlaagmwaldlddfgscfgqdlr1fplna 375
OY 301 VPIYFDNOMVGFDDVESFRTKVSYLKQKGLGAMWALDLDDEFAGFSNQG-RYPLIOT 359
DB 376 lkda1a 381
```

```

OY 360 LRQELS 365

RESULT 3
ENTRY 3 S61550 #type complete
TITLE BRP39 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
ACCESSIONS S61550
REFERENCE S61550
#authors Morrison, B.W.
#submission Submitted to the EMBL Data Library, November 1995
#accession S61550
#status Preliminary
#residues 1-381 #label MOR
#cross-references EMBL:X93035
SUMMARY #length 381 #molecular_weight 43001 #checksum 5107

Query Match 52.1%; Score 1417; DB 14; Length 381;
Best Local Similarity 50.7%; Pred. No. 3.51e-237;
Matches 186; Conservative 84; Mismatches 86; Indels 11; Gaps 9;

DB 23 klvcytswsgyregdgsfda1dp1fclthllysfanssmdlstwmdvlydtlnt 82
OY 2 klvcyftnmaoYRQGEARFLPKDLPSLCTHLIYAFAGMT-NHQSTENDETLYOEFN 60
DB 83 lkrrnplkllsvgvnfgsfgrfkslaasntgsrrtfkxvppflrtfthgfdglawisp 142
OY 61 GKKNPKLTKLTAIGNNEGTOKFTDMVATANNRQTFVNSAIRFLRKSPDGLDMWEXP 120
DB 143 p--r--lr--dkyfstllkelaefkveq-pgrrklllssalssagkva1dygdaqla 196
OY 121 PGSQSAVNDKERFTTLVODLANAFQOEPAQTSKGERLLLSAAVPAQGTYYDAGYEVDK 180
DB 197 hldfslmtydfhgwargtqthsp1frggedasprfndtyavym1rlgapaaklv 256
OY 181 QNIDFVNLMAYDFHGSWEKVTGHNSPLYKKRDESGAASLVNDAVQOMLOKGTIPASKLIL 240
DB 257 mg1ptfsgfllaas-englapisgipgrfkeaglayelcdflkgaevhrlsne 315
OY 241 LGMPTGRSFTLASSSDTRVGAAPATGSGTGPFTKEGMLAYEVCW-KGATKORIQDCK 299
DB 316 kvpfackngvgyehesvkxkvglkxkklagammwaidlddfg-tcqpkeftpln 374
OY 300 KVPYIFRDNOMVGFDDVESFRTKVSYLKQKGLGAMWALDLDDEFAGFSNQG-RYPLIOT 358
DB 375 alkda1a 381
OY 359 TLQELS 365

RESULT 4
ENTRY 4 I48271 #type complete
TITLE BRP39 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
ACCESSIONS I48271
REFERENCE I48271
#authors Morrison, B.W.; Leder, P.
#journal Oncogene (1994) 9:3417-3426
#title neu and ras initiate murine mammary tumors that share genetic
markers generally absent in c-myc and int-2-initiated
tumors.
#cross-references MID:95060797
#accession I48271
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-381 #label RES
#cross-references EMBL:X93035; NID:g1085065; CDS_PID:g1085066
```

## GENETICS

```
#note      gene name brp39
SUMMARY    #length 381 #molecular-weight 43001 #checksum 5107
```

Query Match	52.1%;	Score 1417;	DB 14;	Length 381;
Best Local Similarity	50.7%;	Pred. No. 3.51e-237;		
Matches	186;	Conservative	84;	Mismatches 86;
			Indels 11;	Gaps 9;

Dd 23 klvcyftswgyregvysflpdalqpficthlysfanissdmistwendesnydkln 82  
|||||::||| : || :: |||||::||| :: :: ||| ||||| : :: :  
Qy 2 klvcyfttnaaqyrrogearflprddpslcthliafagmw-nhqsttewndetlyoeen 60

```

Db      83 kIktIntnIktllsvygwkfefekrfseIasnterrtafrsvapIrlsyfIdgIdlawly 142
      || : |||||::|||::| :||| | ||| :|||::| |
Qy     61 GLKKMNPKEKTLTLAGGWNFEQKFIDVAVATANNRQTFNNSAIRLRLKYSIDGDLDMWY 120

```

```
Db      143 p--r-lr--dkqyfstclikelnafctevq-pyreklillsaalsgkvaldtgydaqla 196
          | : ||::| :|| ::||::||::||::||::||::||::||::||::||::||::||::||
Qy     121 PGSGSPAVDKERFTTLVQDLANAFQQEAGTSCGERLLLSAAVPAGQTYYVDAGYEYDCKIA 180
```

Dd 197 ghddflnmlydfbhgwrgtighnspftggqkdrfdrysnnvayavymrlgagaskll 256  
+::+:++++++ + :+++::: + :: ++::: +++:  
Qy 181 QNIDFVNLMAYDFHGSWEKVTGHNSPLYKKRQESGAASLNWDAAVQQWlQGTPASKLI 240

D6   mg|ptfiksitlss -enqjaplsqglpyrftkcaagtlayelcdfkgaevnhsne 315  
      :::||||||| : ::||| : | | ||||| : ||||| : | | : :  
OY 241 LGMPTYGKRFSTLASSSDTRVGAPATGSGTPGPFTKEGMLAYEVCW-KGATKQRIDQ 299

310 KVPYIAFRDQWVGFDEYSEFKTVSYLKQKGLGAMWALDDDDDEAGFSCNQGRY-PLIQ 358  
 300 KVPYIAFRDQWVGFDEYSEFKTVSYLKQKGLGAMWALDDDDDEAGFSCNQGRY-PLIQ 358

DU	5/3	airwaia	301
	::::	!	
QY	359	TLROELS	365

RESULT	5
ENTRY	S27879
TITLE	secretory protein YM-1 precursor - mouse

ACCESSIONS	\$27879
DATE	17-Apr-1993
DESCRIPTION	17-Apr-1993 #sequence_rev1stn 17-Apr-1993 #text_change
	03-May-1996

#authors	#submissions	#description
22/07	Chang, N.C.A.; Liu, C.H.; Chang, A.C.	submitted to the EMBL Data Library, June 1992
		Molecular characterization of a secretory protein (YM-1)

```

transcription factor, macrophages.
#accession S27879
##molecule_type mRNA

```

```
##cross-references EMBL:M94584
CLASSIFICATION  #superfamily Streptomyces chitinase chi40
FEATURE
```

22-399	#product	secretory	protein	YM-1	#status	predicted	#1
		MAT					
SUMMARY	#length	399	#molecular-weight	44528	#checksum	3208	

Query Match	49.48;	Score 1345;	DB 5;	Length 399;
Best Local Similarity	47.78;	Pred. No. 2.04e-223;		
Matches 175; Conservative	87;	Mismatches 102;	Indels 3;	Gaps

D6 lmcyytswardrplegtsfkpgnidpcutllyafagmqneltythgedlrdyeaangl 83  
 :||:|:|:| | : | : ||| ||||| ||||| ||:|:| | | :|:|  
 QY 3 lvcyftnwaqyrqgearflprkdldpdsclthlyafagmtnhqlsttewndetllyqenql 62

```

D5      84 kdkkntelkllaisgwkifgpaisamstwtpgrqllfsgsvlrtrrgyntdglndwgyp 143
      | : : | | | | | | | | | | : : : : | | : : | | | | | | | | | |
OY      63 K-KMNPKLTLIAIGSWNFGIOQFTDMKATANNRQTFFVNSAIRFLPKYSDGDLDEWEY 121

```

Db 144 gsrqspkdkhlfevlvkemrkafeeesvekdiprllltstgagldvksygltrslncis 203

Db 204 lldyqvmtycdlhpckdgytgensplykspdydigsadlnvdsllsywkdhgaasekliv 2633

```

Db      264  gfpayghftflsdpskrtgigabctistpppgkyctdesgllayevctflneatewdaq 3233
| |::|::| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```
Db      324 evpyayqgnwvgydnvrfskkaqlkdnlsgavvwpldmddsfscqhgrhfpltst 363
```

Db 384 lkgdlni 390  
| : : | : :  
391

## RESULT 6

TITLE	ORGANISM	NAME
oviduct-specific glycoprotein 95k precursor - bovine (fragment)	Bos primigenius taurus	#formal_name #common_name cattle
		38-Oct-1995 #accession_number 03-Nov-1995 #start_date

ACCESSIONS	03-May-1996
S57197	
S57197	
REFERENCE	
#authors	
Seedat V : Abe H : Kikuchi M : Satoh T : Uochi H	

#journal	Biol. Reprod. (1994) 50:921-934
#title	Purification and molecular cloning of bovine oviduct-specific glycoprotein.
#accession	S57197

```

#status      preliminary
##molecule_type  mRNA
##residues    1-537  ##label SEN
##cross-references  EMBL: D16639

```

Query Match	48.8%	Score 1328;	DB 14;	Length 537;
Best Local Similarity	51.2%	Pred. No. 3,60e-220:		
SUMMARY	#length 537	#checksum 1095		

```

          matches 107; conservative 74; mismatches 37; inserts 7; gaps
Db 20 klvcyftnwafrpgpaslprldipflcthlvafasmsnqlyrpkdpqdekllypefn 79

```

```

2  ADVATC I AMWQI AVUGSRNE DEVDUDFSDLC I RLLIIE AWMINRQUSI I EMUNDEI - LIGVEFN 50
Db  80 kIkernrgIkllslsgwnfgvwrfttmJstfsmrerfvssvJallrthgfdgJldlffly 1399

```

Db 140 pglspspardtwtfvflleellqafkneagltmrprlllsaaavsgdphvwqkayearlllg 199

```

200 rlldfisvdydhgswekvctghnspfs-l--pgdpks-sa-yamnywrqlgvppkll 2566
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 255 mglptfgrtfnlksasqnelraqavgpsppgkyrtkqagflayyeloccfvrrakkrrwIndg 314  
:::|||||: | : : | | ::::| :||::| |||||: | | :| | |

```

Db      315  yvpyafkgykewygdaisfygkaffikrehfgamwcliddfryfcgtqpfplvht 374
      ||| |::|||:| | |::: |:|||:|:||| | | | | | | | | | | |
300  ttrvrvvrvvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrvrv

```

Db	375	lnn11	379
Or	360	TPort	364

ENTRY	7	146470	#type complete
TITLE		estrogen dependent oviduct protein precursor - sheep	
ORGANISM		#formal_name Ovis orientalis aries, Ovis ammon aries	
DATE		16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Aug-1996	
ACCESSIONS		146470	
REFERENCE		Desouza, M.M.; Murray, M.K. Endocrinology (1995) 136:2485-2496	
#authors		An estrogen-dependent secretory protein, which shares identity with chitinases, is expressed in a temporally and regionally specific manner in the sheep oviduct at the time of fertilization and embryo development.	
#journal			
#title			
#cross-references		EMBL:U01719; NID:g885600; CDS.PID:g885601	
#accession		146470	
#status		preliminary; translated from GB/EMBL/DBJ	
#molecule_type		mRNA	
#residues		1-539 #label DES	
#cross-references		EMBL:U01719; NID:g885600; CDS.PID:g885601	
SUMMARY		#length 539 #molecular_weight 59535 #checksum 2120	
Query Match	48.4%	Score 1318; DB 14; Length 539;	
Best Local Similarity	50.1%	Pred. No. 2,92e-218;	
Matches	183;	Conservative 77; Mismatches 98; Indels 7; Gaps 5;	
Db	23	klyvofnwaafspgsasllprdlpfctthlvafasmannglvpkdpdeklltypefn 82	
Qy	2	klyvofnwaafspgsasllprdlpfctthlvafasmannglvpkdpdeklltypefn 60	
Db	83	klkernrgylkllsvagwntfgtsrftkmlstfnrftkvsiallrhgfogldffly 142	
Qy	61	glkkmnrklklllvgmngtorkftdmvatannrktfynsairpdrkysfpgdlmxy 120	
Db	143	pqlgspardwrtfvllleelgafrneqlltmprlllisaavgdphvlgkaydarllg 202	
Qy	121	pgsogspavndkertrtvlvodlanafqoagotsekerlllssaavpaggqvayagvvdria 180	
Db	203	rllldfsvlsydlhgswekvtnhnsplflspgdkssa-yam--sy--wrglyvpekl 257	
Qy	181	qndefvmlaydfgsmekvtghnsplyrkroesggaafslndavvoowlqgtpasrkl 240	
Db	258	mgllptgrtflhlaasneqlagaaagpasgkytklgaqflayaycsfygtrakrwndg 317	
Qy	241	lgmtygrsflmlasssotrvgaatgsgtppgprtkgcmalayevcsa-kcatkrldio 299	
Db	318	yvpafafgkewwygdalstfykaflktrhefgamwvldlddfngfcgtpfplah 377	
Qy	300	kvrpifpndqmgvddvesfkrvtvslkkgkglgagaywvmlddpafgscnggrfplqt 359	
Db	378	lnnll 382	
Qy	360	lroel 364	
RESULT	8		
ENTRY	A53918	#type complete	
TITLE	chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus sp.)		
ORGANISM	#formal_name Chelonus sp.		
DATE	28-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 03-May-1996		
ACCESSIONS	A53918		
REFERENCE	A53918		
#authors	Krishnan, A.; Nair, P.N.; Jones, D.		
#journal	J. Biol. Chem. (1994) 269:20971-20976		
#title	Isolation, cloning, and characterization of new chitinase stored in active form in chitin-lined venom reservoir.		
#accession	A53918		
#status	preliminary		
#molecule_type	mRNA		

	##residues	1-483	##label	KRI
	##cross-references	GB:U0422		
KEYWORDS	glycosidase; hydrolase			
SUMMARY	#length 483 #molecular-weight 52013	#checksum 5938		
Query Match	43.3%;	Score 1179;	DB 12;	Length 483;
Best Local Similarity	42.5%;	Pred. No. 8,78e-192;		
Matches	158;	Conservative	98;	Mismatches 99;
			Indels	17;
			Gaps	11;
Db	23	kvvvyfgawsvyrrgnkfkdngfdpfcchlylsfvgvngkdvkvdpsldpnlidgf	82	
QY	2	KLVEYFNMAQYRGGEARFLPKDLPSCHLILYAFGMNHOLSTE-WND--ETL--Y	56	
Db	83	gkftsltkhnpvsklmavagvwaagsvpfsgmasdaqtrafaqnvkf1qyqfdgfdi	142	
QY	57	QEFNGLKMMNKLTLLAIGMNGTCGKFPDMMVATANNRQTFVNSAIRFLKRSFSDLDL	116	
Db	143	dweydaqrgsgpadvkmvklckalkkafqg-h---d-yilsaavaapetsaskxydi	195	
QY	117	DMEYPGSGSPFAVDKEREFTTLVLDLANAFQEOATSGKERLLLSAAPACQTYVDAGYEV	176	
Db	196	aemsgyldflnltmyvdthgpdgdtgmhappsasbhsngmeklnvkaaykvyqnvvpk	255	
QY	177	DKIQNDLDFEVLMMYIDHGSWEKTYGINSPLRYKQESGGAASLNVDAVAQWLOKSTPA	236	
Db	256	eklvvgypaygskftlspnsnk1gapavsgagteqpytgenqllgynelcemqkagdev	315	
QY	237	SKLLIGMPTYGRSPTLLASSSDTRRGVATPATSQTPGPTKSGMLAYVEGSM-KGATKQR	295	
Db	316	vqdmekyvpavkqngvsvfdldlaakakagqllkqeglqgmvstetddfkgl-egc-k	373	
QY	296	IQD-QK-VPTYFRNQMVGDFDVSFETKVSXYLKQKGLGAMWALDDDFAGFSCHQGR	353	
Db	374	ypvklainsvlg	385	
QY	354	YPLIQTLRQELS	365	
RESULT	9			
ENTRY	A38221	#type complete		
TITLE	chitinase (EC 3.2.1.14) MFI - nematode (Brugia malayi)			
ORGANISM	chitinae name Brugia malayi			
DATE	04-Mar-1993	#sequence_revision 18-Nov-1994	#text_change	
ACCESSIONS	A38221			
REFERENCE	A38221			
#authors	Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Plessens, W.F.;			
	Perlier, F.B.			
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552			
#title	Transmission-blocking antibodies recognize microfilarial			
	chitinase in Brugian lymphatic filariasis.			
#cross-references	MUID:92179220			
#accession	A38221			
##status	preliminary; not compared with conceptual translation			
##molecule_type	nucleic acid; protein			
##residues	1-504	##label	F0H	
##cross-references	NCBIP:85345			
##note	sequence extracted from NCBI backbone			
KEYWORDS	glycosidase; hydrolase			
SUMMARY	#length 504 #molecular-weight 55971	#checksum 7243		
Query Match	39.9%;	Score 1086;	DB 12;	Length 504;
Best Local Similarity	42.8%;	Pred. No. 3.90e-174;		
Matches	158;	Conservative	94;	Mismatches 102;
			Indels	15;
			Gaps	11;
Db	27	cytunvagyidqegkflpnipnigkchillyafakvdelgdsxpfewndetwskgmys	86	
QY	5	CYFNMAQYRGGEARFLPKDLPSCHLILYAFGMNHOLSTI-EWNDE-T-----LYQ	57	
Db	87	avtklretpnplkvllsygygynfgsalfqglaskqgterifksaalfiknndfgdld	146	
QY	58	EFNGLKMMNKLTLLAIGMNGTCGKFPDMMVATANNRQTFVNSAIRFLKRSFSDGILD	117	





[illegible]

QY	57	OEENLAKMNRKLTLLAIGWNGCTQKFDMDMATANNROTFPVNSAIRLRKRSFDCIDL	116
Db	170	dweyp-addtqatmvl1lkeirseqldayaayq-agp-yhflislaapagehy-sflhm	225
QY	117	DWEYSGSGSPAVDKERFTTLVQDLANAFQEOAQTSKERLLLSAAVPAGQTVYDAGEV	176
Db	226	sdlgqvldvynlmaydyagswsygyhdanlfanspnspp-yntcgaikdy1kqgvpa	284
QY	177	DKIAQNDLFVNLMAVDFGSGWEKTVGHNSPLYKRQESGAASLNDVAAVQOMLOKCTPA	236
Db	285	skivlmpilygraf-estgg-1qgytsgigs-gsw-engikdy-kvlpkagatvgyd	336
QY	237	SKILIGMTYGRSTTLASSSDTRKGPAPGSGTGTGPTTKEGMLATYEVCWKGATKQ--	294
Db	337	svagayydydpskelisfdtrdmnckvskylknlq1gsgmfw	379
QY	295	R1QDQKVPIFRDNQWGFDDVESFKTKVSYLRKQKGGAMVW	337
RESULT	14		
ENTRY		QJ1975	#type complete
TITLE		chitinase (EC 3.2.1.14) 1 precursor - imperfect fungus	
ORGANISM		#formalname Aphanocladium album	
DATE		03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 03-May-1996	
ACCESSIONS		QJ1975	
REFERENCE		QJ1975	
#authors		Blaiseau, P.L.; Lafay, J.F.	
#journal		Gene (1992) 120:243-248	
#title		Primary structure of a chitinase-encoding gene (chl1) from the filamentous fungus Aphanocladium album: similarity to bacterial chitinases.	
#accession		QJ1975	
#molecule_type		rRNA	
##residues		1-423 ##label B1A	
##cross-references		GB:X64104	
COMMENT		This enzyme is essential for the degradation of insoluble chitin.	
GENETICS			
#gene		chl1	
#introns		47/2; 80/2; 97/1	
CLASSIFICATION		#superfamily Streptomyces chitinase chl40	
KEYWORDS		glycosidase; hydrolase	
FEATURE			
1-34		#domain signal sequence #status predicted #label sig\	
35-423		#product chitinase 1 #status predicted #label MAT	
SUMMARY		#length 423 #molecular_weight 46072 #checksum 8802	
Query Match		21.2%; Score 576; DB 5; Length 423;	
Best Local Similarity		36.2%; Pred. No. 5,50e-79;	
Matches 101; Conservative		65; Mismatches 95; Indels 18; Gaps 16;	
Db	115	lkkgrmmkvnmlsigytwst-nrpaassatctkttagavgmkgwfdgdldideyp	173
QY	62	LKKMPKTKTLTLAGGWFQGTFTDMVATVNNRQSTVNSAIRLRKYSFDDGLDDEYR	121
Db	174	-adatqagmvl1lqavsejdsyaayq-akg-bhflislaapaghyoklka-e-1g	228
QY	122	GSQGSPPAYDKERFTTLVQDLANAFQEOAQTSKERLLLSAAVPAG-CTYDAGEVDKIA	180
Db	229	kvldylnlmaydyagswnygyhdanlyanpnpnatp-yntdavgay1ngvpankiv	287
QY	181	QNLDFVNLMAVDFHGSWEKTVGHNSPLYKRQESGAASLNDVAAVQOMLOKGPASKLI	240
Db	288	lqmpilygrsfq-qt-e-gg-1qkpyngjgs-gsw-engikdykalpk-agatvkcdttak	339
QY	241	IGMPYGRSFTLASSSDTRKGPAPGSGTGTGPTTKEGMLATYEVCWKGATKQRIODOK	300
Db	340	gcysydpstkelisfdtrdmnckvskylkq1gsgmfw	378
QY	301	VPIR-RD-NQWGFDDVESFKTKVSYLRKQKGGAMVW	337





\*\*\*\*\*  
 M O N J U N 3 0 0 8 : 5 9 : 1 7 1 9 9 7  
 \*\*\*\*\*  
 (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by IntelliGenetics, Inc.

Mparch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Jun 27 16:59:40 1997; Maspar time 12.05 Seconds  
 Tabular output not generated. 656,435 Million cell updates/sec

Title: >US-08-663-618A-14  
 Description: (1-373) from US08663618A.pep  
 Perfect Score: 2721  
 Sequence: 1 AKLVCFYFTNMAQYRQGEARF.....YPLIQTURQELSLPELPDGT 373

Scoring table:  
 PAM 150  
 Gap 11

Searched: 59021 segs, 21210388 residues

Post-processing: Minimum Match 08  
 Listing first 45 summaries

Database:

swiss-prot34  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11

Statistics: Mean 48.978; Variance 85.387; scale 0.574

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1488	54.7	383	4	GP39_HUMAN CARTILAGE GLYCOPROTEI	0.00e+00
2	1086	39.9	504	2	CHIT_BRUMA ENDOCHITINASE PRECURS	6.23e-219
3	1015	37.3	554	2	CHIT_MANSE ENDOCHITINASE PRECURS	4.86e-202
4	698	25.7	689	2	CHIT_BACCI CHITINASE A1 PRECURSO	1.82e-127
5	576	21.2	423	2	CHIT_APRAL CHITINASE 1 PRECURSOR	2.50e-99
6	561	20.6	423	2	CHIT_TRIHA 42 KD ENDOCHITINASE P	6.68e-96
7	546	20.1	427	2	CHIA_SERMA ENDOCHITINASE 1 PRECU	1.75e-92
8	437	16.1	563	2	CHIA_SERMA CHITINASE A PRECURSOR	6.12e-68
9	395	14.5	551	2	CHIT_NPVAC PROBABLE ENDOCHITINAS	1.16e-58
10	345	12.7	820	2	CHIA_ALUSO CHITINASE A PRECURSOR	8.67e-48
11	337	12.4	499	2	CHIB_SERMA CHITINASE B PRECURSOR	4.53e-46
12	301	11.1	619	2	CHIT_STRLI CHITINASE C PRECURSOR	2.01e-38
13	287	10.5	610	2	CHIT_STRLI CHITINASE 63 PRECURSO	1.72e-35
14	263	9.7	1146	5	KTXA_KLULA KILLER TOXIN ALPHA AN	1.56e-30
15	172	6.3	36	4	GP39_BOVIN CARTILAGE GLYCOPROTEI	8.27e-13
16	172	6.3	83	3	EDOS_PAPAN ESTRADIOL-DEPENDENT O	8.27e-13
17	121	4.4	597	2	CHIX_STROI EXOCHITINASE 1 PRECUR	3.59e-04
18	108	4.0	452	4	HEME_RHOSH COPROPORHYRINOEN II	3.08e-02
19	110	4.0	497	2	CHID_BACCI CHITINASE D PRECURSOR	1.59e-02
20	104	3.8	422	10	UMUC_ECOLI UMUC PROTEIN.	1.13e-01
21	100	3.7	364	11	YBR5_YEAST HYPOTHETICAL 40.9 KD	3.99e-01
22	101	3.7	1045	7	PRTT_SERMA EXTRACELLULAR SERINE	2.92e-01

RESULT ID	GP39_HUMAN	STANDARD	PRT	383 AA.	ALIGNMENTS
AC	P36222; P30923;				
DT	01-JUL-1993 (REL. 26, CREATED)				
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)				
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)				
DE	CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
NC	EUTHERIA; PRIMATES.				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE-CARTILAGE;				
RX	MEDLINE; 94064658.				
RA	HAKALA B.E., WHITE C., RECKLIES A.D.;				
RL	J. BIOL. CHEM. 268:25803-25810(1993).				
RN	[2]				
RP	SEQUENCE OF 22-45.				
RX	MEDLINE; 90328983.				
RA	NYIKROS P., GOLDS E.E.;				
RL	BIOCHEM. J. 269:265-268(1990).				
CC	- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO				
CC	RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.				
CC	- TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL				
CC	CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,				
CC	PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.				
CC	- PTM: GLYCOSYLATED.				
CC	- SUBCELLULAR LOCATION: EXTRACELLULAR.				
CC	- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.				
DR	EMBL; M80927; G348912; -				
DR	PIR; S10677; S10677.				
DR	PIR; A33162; A33162.				
DR	PROSITE; PS01095; CHITINASE_18.				
KW	GLYCOPROTEIN; SIGNAL.				
FT	SIGNAL 1 21				
FT	CHAIN 22 383				
FT	CARBOHYD 60 60				
SO	SEQUENCE 383 AA; 42613 MW; D0A5280F CRC2;				
Query Match	54.7%	Score 1488;	DB 4;	Length 383;	
Best Local Similarity	53.3%	Pred. No. 0.00e+00;			
Matches 195;	Conservative 83;	Mismatches 79;	Indels 9;	Gaps 5;	
DB	23	AKLVCFYFTNMAQYRQGEARF	LPKDLPSLCTHLIYARAGMTNQLSTEWNDLTVQEPNG	61	

```

Db      83 lknpmllkllsvgyvmfsgqrtsklsasntqsr:rttkvfyvplrlfthgldldlwlyp 142
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      62 LKKNMPLKLTLLLAGNMFGTQKTTDMWATLNNQKTVNSLRFRKRYSPFDGLDMEYR 121
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      143 grt-----dkqhtftllkkmkaefikeaq-pykkqjllsaalsagkvrtldssydlahtsq 196
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      122 GSGGSPANVDKRFPTTIVODLANARQGEAQISGKRRLLLSAAPVARGRYVDAGYEVDKIAQ 181
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      197 hldflslmcydflnqawrgtcghnspilfriggcdaspdf:tsntdyavgymlllgaspaklvm 256
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      182 NLDEVNLMAYADYFHSGSWEKVTGSHNSPLTKRQEESSAASLNVDAVAQWQLMGKTPASRLTL 241
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      257 giprfqfsftlass-eyvgagapisgpqalpgrftkeagtlaiyeicdlilgratvhrtlrgq 315
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      242 GMPYRGREFTLASSSDPRVGAAPAGSGSPGPFTTEGGMLAYEYECVSR-KGATRIORIDOK 300
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      316 vpyatkgwqwgvyddgsevsykayllldrqldqagmawmallddvgqsfcgqdlrfrptlna 375
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      301 VPLYFRDQWQVGFDDVESFRTKVSYLKQGGGAMVWALDDDFAGFSCKNOG-RYPLIQT 359
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      376 lkdaa 381
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      360 LRQELS 365
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT	2	STANDARD;	PTI;	504 AA.
ID	CHIT. BRUA			
AC	P29030;			
DT	01-DEC-1992 (REL. 24, CREATED)			
DT	01-FEB-1992 (REL. 24, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MFI ANTIGEN).			
OS	BRUGIA MALAYI.			
OC	EUBARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; SPIRURIDA.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE; 92179220.			
RA	FURHRMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.,			
RL	POURMAN, N.A., ACAD. SCI. U.S.A. 89:1548-1552(1992).			
CC	-1- FUNCTION: THE MFI ANTIGEN IS A MICROFILARIAL CHITINASE, WHICH MAY			
CC	FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-			
CC	FILARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND			
CC	TRANSMISSION.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF			
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-1- PTM: O-GLYCOSYLATED.			
CC	-1- KNOWN TO BIND CALCIUM.			
CC	-1- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MFI ANTIGEN CORRESPOND			
CC	WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.			
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES).			
DR	EMBL; M73689; G156064; -.			
DR	PIR; A38221; A38221.			
DR	PROSITE; PS01095; CHITINASE.18.			
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;			
KW	REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.			
FT	SIGNAL	1	22	
FT	CHAIN	23	504	ENDOCHITINASE.
FT	DOMAIN	23	400	CATALYTIC.
FT	DOMAIN	401	450	SER/THR-RICH (LINKER).
FT	DOMAIN	407	448	3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT	ACT_SITE	148	148	PROTON DONOR (BY SIMILARITY).
SO	SEQUENCE	504 AA;	55971 MM;	4DAY5EB CRC32;
Query Match		39.9%;	Score 1086;	DB 2; Length 504;
Best Local Similarity		42.8%;	Ident. No. 6,22e-219;	
Matches 18;	Conservative		94;	Mismatches 102; Indels 15; Gaps

Db	87	avtkhrenepljxvllisgygnfesaftibxaksagctfeifisafafikmndofrdld	146
Qy	58	EFNGIKKMKPKLTKTLTLAGGWNSTQKFTDMVAANRRQFFVSALEFKKYSFDDGLD	117
Db	147	wepv-v-gv-a-e-eh-aklveamktafveeaktsqkrllltaavsaagkttdgsyve	200
Qy	118	WEYGSOGSPAVDKERRTTLVQDANNAFQGEAQSCKGERLLLSAAYPAGTYYDAGYEVD	177
Db	201	slghnfdllfmsydlhgsweknydlhghk1hptkgevsglqifntefaadyasakgmpe	260
Qy	178	KIAONLDFVNLMADYFHSWEKVTGHNSSPLYRKQEESSGAASLNVDAAYQOMLQKTPAS	237
Db	261	klllgilpmayagvcladmpsetajgaaasrpsasaaklnpagtfsayweickylkegyketv	320
Qy	238	KLIIGMPYKGSFTLASSSDTRVAPATGCTGTPPFPFKEGGMALAYEVC5M-KGATKRI	296
Db	321	hgqvgagmvmvgddvydyndneetrlkmkwlkhegygsafialdffdffgkscgkyyp	380
Qy	297	QDQKV-PIEFHDNOMWGFDDVESEKTVSYLKOKGLGAGAAWMAALDDDFAGFCNCGRYP	355
Db	381	llnaisel 389	
Qy	356	LIQTLROEL 364	

ID	RESULT	3	STANDARD;	PRT;	554 AA.
AC	P36362;				
DT	01-JUN-1994	(REL. 29, CREATED)			
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995	(REL. 31, LAST ANNOTATION UPDATE)			
DE	ENOCCHITINASE PRECURSOR	(EC 3.2.1.14).			
OS	MANDUCA SEXTA (TOBACCO HAWKMOTH) (TOBACCO HOMINORM).				
CC	EPICAROTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 93357793.				
RA	KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;				
RL	INSECT BIOCHEM. MOL. BIOL. 23:691-701(1993).				
CC	-1- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING PROCESS.				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF				
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.				
CC	-1- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,				
CC	BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH				
CC	INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A				
CC	RAPID DECLINE IS SEEN. IN THE GUT IT IS DETECTED ON DAY 6 WITH LOWER				
CC	LEVELS SEEN ON DAYS 0, 7 AND 8.				
CC	-1- TISSUE SPECIFICITY: EPIDERMIS AND GUT.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
CC	HYDROLASIS).				
DR	EMBL; U02270; G406049; -;				
DR	EMBL; S64757; -; NOT_ANNOTATED.CDS.				
DR	HSSP; P07254; IGTN.				
DR	PROSITE; PS01095; CHITINASE_18.				
RR	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.				
FT	SIGNAL	1	19	POTENTIAL.	
FT	CHAIN	20	554	ENDOCHITINASE.	
FT	DOMAIN	396	453	SER/THR-RICH.	
FT	ACT_SITE	146	146	PROTON DONOR (BY SIMILARITY).	
FT	CARBOHYD	85	85	POTENTIAL.	
FT	CARBOHYD	303	303	POTENTIAL.	
FT	CARBOHYD	407	407	POTENTIAL.	
FT	CARBOHYD	545	545	POTENTIAL.	
QO	SEQUENCE	554 AA;	62203 MW;	FA87F8AD CRC32;	

[illegible]



181 QNDFVNLMAYPDHGSMKEVYTGHNSPLYKROESGAALNDVAAYQVMTQKGPASKL 240  
Db 288 lmplygrsfq-gt--eg-igkpyngfqs-gsw--engivdykalpk-asatkcddtak 339  
QY 241 LGMPTGRSTTLASSSDTRVGAPATGSGTGPFTKEGMLAYEVCWSMGATKORIODOK 300  
Db 340 gcysydpstkelisfdtpaistekvsklgkqigntmfw 378  
QY 301 VPYIF-RD-NQWGFDDVESFKTKVSYLKOKGLGAMVW 337

RESULT 6  
ID CHIA-TRIHA STANDARD; PRT; 423 AA.  
AC P48827;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE 42 KD ENDOCHITINASE PRECURSOR (EC 3.2.1.14).  
GN CHIT42.  
OS TRICHODERMA HARZIANUM.  
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.  
RX MEDLINE; 95269313.  
RA GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LOBELLE A.,  
RA PINTOR-TORO J.A.;  
RL CURR. GENET. 27:83-89(1994).  
CC -1- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH/CELL DIVISION  
AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL  
AGENT.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE  
REPRESSOR.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
HYDROLASES).  
CC EMBL; 578423; G999376; -.  
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;  
KW CHITIN-BINDING.  
FT SIGNAL 1 22 POTENTIAL.  
FT PROPEP 23 34  
FT CHAIN 35 423 42 KD ENDOCHITINASE.  
FT ACT\_SITE 171 171 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 218 218 POTENTIAL.  
SQ SEQUENCE 423 AA; 46056 MW; 899DA50A CRC32;

Query Match 20.6%; Score 561; DB 2; Length 423;  
Best Local Similarity 35.7%; Pred. No. 6,68e-96;  
Matches 101; Conservative 71; Mismatches 94; Indels 17; Gaps 13;

Db 111 qltk-vtkarqglkvlisigvwtst-nfpaastadnknfaktaitfkdvgfgdtd 168  
QY 57 QEENGKAKMPPKLTLLAIGMNGFTQKFTDMATANNROTFTVNSAIRFKRSFDDLD 116  
Db 169 dweyp-adatqasamlllkevsqrdayaagy-apy-yhflitlaapagkdy-sklr 224  
QY 117 DWIYPSGSGPAYDKERTTLVODLANAFQOEQOTSKEKELLSAAYPAQOTVVDAGYEV 176  
Db 225 adigvldylnlnaydagsfsltgndanlfnpsnpatp-fntdsavkdyingvpa 283  
QY 177 DKLAQNLDVFNLMAYDPHGSWEKVTGHNPSLYKROESGAALNDVAAYQVMTQKOTPA 236  
Db 284 nkivlmplygrsf--qntag--igqlyngvgs-gsw--eagivdykalpk-agatvqd 335  
QY 237 SKLILGMPYGRSFTTLASSSDTRVGAPATGSGTGPFTKEGMLAYEVCWSMGATKORI 296  
Db 336 svakgyysnakkelisfdtpminkvaylsigvgsmfw 378  
QY 297 QDQKVPYIFRD-NQWGFDDVESFKTKVSYLKOKGLGAMVW 337

RESULT 7  
ID CHIL-COCIM STANDARD; PRT; 427 AA.  
AC P54196;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE ENDOCHITINASE 1 PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)  
DE (CF-ANTIGEN) (CF-AG).  
GN CTS1.  
OS COCCIDIOIDES IMMITIS.  
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-C735;  
RX MEDLINE; 96144270.  
RA PISHKO E.J., KIRKLAND T.N., COLE G.T.;  
RL GENE 167:173-177(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SILVEIRA;  
RA SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
HYDROLASES).  
DR EMBL; 141663; G1200190; -.  
DR EMBL; U51271; G1256769; -.  
DR EMBL; U33265; G1255728; -.  
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;  
KW GLYCOPROTEIN.  
FT SIGNAL 1 2 POTENTIAL.  
FT CHAIN 3 427 ENDOCHITINASE 1.  
FT CARBOHYD 387 387 POTENTIAL.  
FT CONFLICT 15 47 RWLSRLCLELCEGRMFLSTVATVATVDIQ -> VQAS  
FT CONFLICT 159 199 SMSMNPYVPEAEPEGGRSVYFVW (IN REF. 2).  
SQ SEQUENCE 427 AA; 47629 MW; 6CB9AF73 CRC32;

Query Match 20.1%; Score 546; DB 2; Length 427;  
Best Local Similarity 34.1%; Pred. No. 1.75e-92;  
Matches 95; Conservative 71; Mismatches 94; Indels 19; Gaps 17;

Db 115 lkknnnllkllisigvwt-spnfxkpasteegrkfadtislkmdldfgddidwep 173  
QY 62 LKMANRKLTLAIGMNGFTQKFTDMATANNROTFTVNSAIRFKRSFDDLDWEP 121  
Db 174 ede-kqandfvlllkacrealdayak-hpngk-kflltlaapagpny-nk-1klaemd 228  
QY 122 GSGSPAYDKERTTLVODLANAFQOEQOTSKEKELLSAAYPAG-QTVVDAGYEVDKIA 180  
Db 229 kyldfnlnaydagsvdkvsglmsnvfptctkp-estfssdkavkdyikagvpankiv 287  
QY 181 QNDFVNLMAYPDHGSMKEVYTGHNPSLYKROESGAALNDVAAYQVMTQKGPASKL 240  
Db 288 lmplygrsf--ast-dg-igtsfngvg-gsw--engivdykalpk-qgaqvteledia 339  
QY 241 LGMPTGRSFTTLASSSDTRVGAPATGSGTGPFTKEGMLAYEVCWSMGATKORIODOK 300  
Db 340 asysydknkryllsydtkvlagkkaeyilkngmggm-w 377  
QY 301 VPYIF-RDQW-VGFDDVESFKTKVSYLKOKGLGAMVW 337

RESULT 8  
ID CHIA-SERMA STANDARD; PRT; 563 AA.  
AC P07254;  
DT 01-APR-1988 (REL. 07, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE A PRECURSOR (EC 3.2.1.14).  
GN CHIA.  
OS SERRATIA MARCESCENS.



OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC ENTEROBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KOO J.C., LIM C.O., CHOI Y.J., KIM C.Y., BARK J.D., LEE S.Y.,  
RA KOO M.J.;  
RL SUBMITTED (JAN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 990 / QMB1466;  
RA JONES J.D.G., GRADY K.L., SUSLOW T.V., BEDBROOK J.R.;  
RL EMO J. 5:467-473(1986).  
RN [3]  
RP REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE; 95219379.  
RA PERAKIS A., TEMS I., DAUTER Z., OPPENHEIM A.B., CHET I., WILSON K.S.,  
RA VORGIAS C.E.;  
RL STRUCTURE 2:1169-1180(1994).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL; L01455; G152818; -.  
DR EMBL; X03657; G46831; -.  
DR PIR; A25090; A25090.  
DR HSSP; P07254; ICTN.  
DR PROSITE; PS01095; CHITINASE.18.  
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL.  
FT SIGNAL 1 23  
FT CHAIN 24 563 CHITINASE A.  
FT DOMAIN 150 563 CATALYTIC.  
FT ACT\_SITE 315 563 PROTON DONOR (PROBABLE).  
FT ACT\_SITE 391 563 PROBABLE.  
FT CONFLICT 76 77 TA -> GP (IN REF. 2).  
FT CONFLICT 139 139 A -> P (IN REF. 2).  
FT CONFLICT 226 226 V -> I (IN REF. 2).  
FT CONFLICT 395 395 P -> A (IN REF. 1 AND 2).  
FT CONFLICT 410 430 PAKKPTATTVNGVNALAQ -> RPSGRRLHGHGRRQC  
FT CONFLICT 437 437 AAQO (IN REF. 1 AND 2).  
FT CONFLICT 464 467 V -> I (IN REF. 2).  
FT CONFLICT 473 473 ATGP -> HRA (IN REF. 2).  
FT CONFLICT 484 484 K -> E (IN REF. 2).  
FT CONFLICT 484 484 G -> S (IN REF. 2).  
SQ SEQUENCE 563 AA; 60979 MW; 7764DB57 CRC32;  
Query Match 16.1%; Score 437; DB 2; Length 563;  
Best Local Similarity 28.8%; Pred. No. 6,12e-68;  
Matches 93; Conservative 84; Mismatches 123; Indels 23; Gaps 19;  
Db 231 pfalaiaqkqyrtavddpykgnfgqlmaikqahpdkilpslsgwtlsdpff-mgdkvk 289  
QY 36 AFAGMNHDLSTTEWDE-T-LYDFNGKLKKNPKIKTLALIGMNGFTOKCFDMVATAN 93  
Db 290 -rdifsgvkefqtqkffdgvdldwefpggkgnpnlsgpdgeeyvlmlkelrmlq 348  
QY 94 NROTFVNSAIRFLKRYSF-DGLDLDWEYPSQGS-PAV---DKEFTTLVODLANAFQO 147  
Db 349 laaetgr-kyelitsaiaqdkldkavynvaq--nsmdhiflmsydfgpfdkniqht 405  
QY 148 EAQTSKERLLLSAAYPAQOTYVD-AGYEVDKAQNLDEYNLMAYPHGSWE-KVYGHNS 205  
Db 406 alnapakpdtayt-tvng-vnallagvykpgkvvvgtamyrgrtgvngynqnlpfct 463  
QY 206 PLYKROEEGGAASLNVDAVQOMLOKGTSPASKLILGMPYGRSFTLASSSDTRVGAPAT 265  
Db 464 atg-pvkgvkwngivdgrtlaqfmsgwytydataaapvyfktstgdlitddarsq 532  
QY 266 GSGTPEPFKEGGMALYEVCS-WKGAIRQ-RIQD-QKVPYIR-DN-QWGFDDVESR 320  
Db 523 akgyvldkqglgflsweidahn 545  
QY 321 TKVSYLKQKGLGGMVWALDLD 343

RESULT 9  
ID CHIT\_NPVAC STANDARD; PRT; 551 AA.  
AC P41684;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).  
OS AUTOGRAHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACNPV).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C6.  
RX MEDLINE; 94303173.  
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;  
RL VIROLOGY 202:586-605(1994).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL; L22858; G559195; -.  
DR PROSITE; PS00014; ER-TARGET.  
DR PROSITE; PS01095; CHITINASE.18.  
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.  
FT SIGNAL 1 17  
FT CHAIN 18 551 POTENTIAL.  
FT ACT\_SITE 305 551 PROBABLE ENDOCHITINASE.  
FT CARBOHYD 173 551 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 444 551 POTENTIAL.  
FT SITE 548 551 POTENTIAL.  
SQ SEQUENCE 551 AA; 61368 MW; BDB124DB CRC32;  
Query Match 14.5%; Score 395; DB 2; Length 551;  
Best Local Similarity 29.2%; Pred. No. 1,16e-58;  
Matches 95; Conservative 80; Mismatches 122; Indels 28; Gaps 27;  
Db 221 pwaavqkpgkysawn-ep-ykgnfgqlmaaklanphkllpslsgwtlsdp-fyfmhcv 277  
QY 36 AFAGMNHDLSTTEWDETLYO-EFNGIK--KM-NPKIKTLALIGMNGFTOKCFDMVAT 91  
Db 278 -ekinvfsvvefivqwmffdgvdldwefpggkgnpnlsgdgdadgdky111leelrm 336  
QY 92 ANNRQTFVNSAIRFLKRYSF-DGLDLDWEYPSQGS-PAV-D-K-E-R-FTTLVODLANA 144  
Db 337 lddleagt-grv-yelitsaiaqdklavvnya-e-agkslqkifmsydfkgsantdl 392  
QY 145 EAGTSGKERLLLSAAYPAQ-QTYYDAGYEVDKINQNLDPVNLMAIDFGSMKVT- 201  
Db 393 gyqtlvya-ps-wnseelythvavdallkqgvdpnkliivgamygrvtytnyndy 450  
QY 202 GHSNPLYKROEEGGAASLNVDAVQOMLOKGTSPASKLILGMPYGRSFTLASSSDTRVG 261  
Db 451 fsgtng-psagtwedgyvdyrtdqldlnnyyrtfdaaqaayvtdksqgdlstfsvds 509  
QY 262 APTGSGTPEPFKEGGMALYEVCSWKGAIRQIRIQD-KVPYIR-RDN-QWGFDDVES 318  
Db 510 vlgkvkyvdrnkllgflfaweidahn 534  
QY 319 FKTVSYLKQKGLGGMVWALDLD 343  
RESULT 10  
ID CHIA\_ALTSO STANDARD; PRT; 820 AA.  
AC P32823;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE A PRECURSOR (EC 3.2.1.14) (CH1-A).  
GN CH1A.  
OS ALTERNOMAS SP. (STRAIN O-7).  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC VIBRONACEAE.  
RN [1]



Db 425 gsdgsk-idaa-dygeaskyidkwnmtygfifgawak-ngptaphnplitaydglpqgfn 481  
QY 163 vpgqotvvdgyevdkiqiaonldfvnlmavdphswnvntshnsplykkrqesgaaa-sln 221  
Db 482 tadamakfkksqyacklllgfygrgwt-gvtqsap-ggtatgpat-gty--eaagied 536  
QY 222 vdaavqomlwkgtfaskllilgmptlgrsftllasssdtrvnapatgsgtppftrksgmla 281  
Db 537 y-kvlnsapatgltag--tayahcgsnwnsydtpatlkxmdwaegqglggaftweifsg 593  
QY 282 yvevscwmkcatkroridokvpyifrdnqmwgfpdvesfktkvsylkqkglgamvwaldl 341  
Db 594 dtang 598  
QY 342 ddfag 346

RESULT 13  
ID CHIT\_STRPL STANDARD: PRT: 610 AA.  
AC P1120;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE CHITINASE 63 PRECURSOR (EC 3.2.1.14).  
GN CHTA  
OS STREPTOMYCES PLECATUS.  
OC PROMARYOTA; FRIMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92192480.  
RA ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;  
RL GENE 111:69-76(1992).  
RN (2)  
RP SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.  
RX MEDLINE: 88087127.  
RA ROBBINS P.W., ALBRIGHT C., BENFIELD B.;  
RL J. BIOL. CHEM. 263:443-447(1988).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC -1- INDUCTION: BY CHITIN.  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE IIT-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
CC (CBD).  
DR EMBL: M82804; G153216; -;  
DR EMBL: M18397; G153209; -;  
DR PIR: A29912; A29912.  
DR PIR: JH0573; JH0573.  
DR HSSP: P07234; 1CTN.  
DR PROSITE: PS00018; EF\_HAND.  
DR PROSITE: PS00561; CBD\_BACTERIAL.  
DR PROSITE: PS01095; CHITINASE\_18.  
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.  
FT SIGNAL 1 30  
FT CHAIN 31 610 CHITINASE 63.  
FT DOMAIN 35 140 CELLULOSE-BINDING.  
FT DOMAIN 148 229 FIBRONECTIN TYPE-III.  
FT DOMAIN 236 610 CATALYTIC.  
FT ACT\_SITE 383 383 PROTON DONOR (BY SIMILARITY).  
FT CONFLICT 3 3 F -> I (IN REF. 2).  
SQ SEQUENCE 610 AA: 63974 MW: 2F5E8E35 CRC32;

Query Match 10.5%; Score 287; DB 2; Length 610;  
Best Local Similarity 26.4%; Pred. No. 1,72e-35;  
Matches 78; Conservative 63; Mismatches 134; Indels 21; Gaps 16;

Db 320 fnglrmkkaeyphklllysfgrwtwsgg-fpdavknpaafakschdivedprwadvdq1 378  
QY 56 yofenglkkknpkrlklllaigwnwfgokftdmvafannqotfvenslrlrkys-fdgl 114  
Db 379 dlwepnaeglscedesapnatsmmkamraefgqdyllta-a---vtadsgdgsk-id 433

QY 115 dlldmefpssgg-s--pavdkerttvlvodiannfocbaotskgerlllsaaavagottyd 171  
Db 434 aa-dygeaskyidkwnmtydfffagawak-ngptaphnplnaydglpqgfttadamakfk 491  
QY 172 agvevdkiqiaonldfvnlmavdfigswewkytghnsplykkrqesgaaa-slnvdaavqoml 230  
Db 492 skvpadklllgfygrgwt-gvtqsap-ggtatgpa-gty--eaagiedy-kvlnsc 545  
QY 231 oktpaskllilgmptlgrsftllasssdtrvnapatgsgtppftrksgmlaayvevscwmg 290  
Db 546 patgtag--tayahcgsnwnsydtpatlkxmdwaegqglggaftweifsgdtng 599  
QY 291 atkroridokvpyifrdnqmwgfpdvesfktkvsylkqkglgamvwaldddfag 346

RESULT 14  
ID KTXA\_KLULA STANDARD: PRT: 1146 AA.  
AC P09805;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)  
DE (ENOCCHITINASE (EC 3.2.1.14)).  
OS KLUYVEROMYCES LACTIS (YEAST).  
OG PLASMID PGK1-1.  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CBS 2359/152;  
RA SOR F., FUKUHARA H.;  
RL CORR. GENET. 9:147-155(1985).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84297209.  
RA STARK M.J.R., MILEHAM A.J., ROMANOS M.A., BOYD A.;  
RL NUCLEIC ACIDS RES. 12:6011-6030(1984).  
RN (3)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85037931.  
RA TOKUNAGA M., KAMAMURA A., HISHINUMA F.;  
RL NUCLEIC ACIDS RES. 12:7581-7597(1984).  
RN (4)  
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.  
RX MEDLINE: 87004569.  
RA STARK M.J.R., BOYD A.;  
RL EMBO J. 5:1995-2002(1986).  
RN (5)  
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.  
RX MEDLINE: 90259069.  
RA BRADSHAW H.D. JR.;  
RL NATURE 345:299-299(1990).  
RN (6)  
RP CHITINASE ACTIVITY OF ALPHA-SUBUNIT.  
RX MEDLINE: 91301161.  
RA BUTLER A.R., O'DONNELL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;  
RL EUR. J. BIOCHEM. 199:483-488(1991).  
CC -1- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE, ALONG WITH  
CC THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE  
CC TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE  
CC TOXIN) TO GAIN ENTRY INTO THE CELL.  
CC -1- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-  
CC SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.  
CC -1- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,  
CC BETA AND GAMMA.  
CC -1- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.  
DR EMBL: X07127; G28429; -;  
DR EMBL: X00762; G28449; -;  
DR EMBL: X01095; G28499; -;  
DR PIR: S07915; S07915.  
DR HSSP: P02877; 1HEV.  
DR PROSITE: PS00026; CHITIN\_BINDING.



\*\*\*\*\*  
MURINE  
(TM)  
\*\*\*\*\*

Release 2.1d John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligentics, Inc.

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 27 17:08:30 1997; Maspar time 9.60 Seconds  
423.400 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-663-618A-15  
Description: (1-373) from US08663618A.pep  
Perfect Score: 2717  
Sequence: 1 ALVCFYFTNMAQYRGEARF.....YPLIOTRQELSPYLSSGT 373

Scoring table:  
PAM 150  
Gap 11

Searched: 92623 segs, 10896596 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq26  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19

Statistics: Mean 34.957; Variance 153.180; scale 0.228

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1328	48.9	537	14 R73991	Bovine oviduct specif	9.77e-114
2	1303	48.0	668	14 R73993	Hamster oviduct specif	2.57e-111
3	1303	48.0	718	14 R73992	Murine oviduct specif	2.57e-111
4	881	32.4	321	14 P81342	Polypeptide involved	1.13e-70
5	597	22.0	371	13 R70025	Tobacco chitinase enc	1.07e-43
6	597	22.0	371	13 R70029	Tobacco cluster-A pro	1.07e-43
7	581	21.4	424	12 R63441	Trichoderma harzianum	3.42e-42
8	576	21.2	389	10 R56861	Aphanocladium album m	1.01e-41
9	576	21.2	423	6 R33072	Aphanocladium album p	1.01e-41
10	569	20.9	389	6 R33068	Aphanocladium album m	4.58e-41
11	463	17.0	866	19 W02159	Soluble chitinase.	3.52e-31
12	389	14.3	561	12 R48823	Chitinase A.	2.29e-24
13	372	13.7	561	9 R48981	Bacterial chitinase-A	8.14e-23
14	271	10.0	238	6 R32547	HANPV IE-1 gene prote	9.44e-14
15	219	8.1	1046	19 W02156	Periplasmic chitodext	3.12e-09
16	139	5.1	25	13 R70751	YKL-40 N-terminal pep	1.22e-02
17	139	5.1	25	11 R70745	YKL-40 N-terminal seq	1.22e-02
18	136	5.0	29	11 R60073	Bovine Oviduct Glycop	2.09e-02
19	101	3.7	1070	4 R21521	Alpha galactosidase.	8.48e+00
20	99	3.6	398	18 R95856	S. pyogenes speb gene	1.17e+01

21	99	3.6	428	16 R86905	Murine BMP-9.	1.17e+01
22	99	3.6	553	14 R31372	Murine bone morphogen	1.17e+01
23	99	3.6	553	14 R80445	Murine soluble epoxid	1.17e+01
24	99	3.6	958	14 P81181	Sequence of glucosamyl	1.17e+01
25	99	3.6	958	2 R07575	Glucosamylase encoded	1.17e+01
26	95	3.5	438	9 R47259	Pre-pro-VGR1.	2.23e+01
27	92	3.4	1356	5 R26989	Novel type III RTK en	3.58e+01
28	91	3.3	19	13 R70752	YKL-40 internal pepti	4.18e+01
29	91	3.3	19	12 R70746	YKL-40 internal pepti	4.18e+01
30	90	3.3	256	17 R97235	Mutu putative oncogen	4.89e+01
31	90	3.3	459	17 R88058	Protein encoded by Mu	4.89e+01
32	90	3.3	575	3 P50018	Sequence encoded by t	4.89e+01
33	90	3.3	709	3 P70423	FIXD nitrogen fixatio	4.89e+01
34	90	3.3	709	3 P50017	Sequence encoded by t	4.89e+01
35	87	3.2	513	9 R98125	Human lymphocyte cell	6.66e+01
36	88	3.2	513	9 R47264	Pre-pro BMP6.	6.66e+01
37	88	3.2	513	2 R07311	Human Bone Morphogene	6.66e+01
38	88	3.2	513	6 R32904	Human BMP-6 peptide s	6.66e+01
39	88	3.2	542	16 R36735	Human BMP-6.	6.66e+01
40	87	3.2	896	17 R90772	Bacillus stearothermo	7.77e+01
41	88	3.2	896	17 R92750	Human EGF receptor su	6.66e+01
42	86	3.2	2332	8 R43257	Human Factor VIII.	9.05e+01
43	87	3.2	2351	3 P50319	Human antithaemophilic	7.77e+01
44	86	3.2	2351	3 P50059	Human factor VIII.	9.05e+01
45	86	3.2	2351	13 R78223	Human Factor VIII:C.	9.05e+01

## ALIGNMENTS

RESULT 1  
ID R73991 standard: Protein; 537 AA.  
AC R73991;  
DT 10-JAN-1996 (first entry)  
DE Bovine oviduct specific glycoprotein.  
KM Bovine oviduct specific glycoprotein; recombinant production;  
KW BOGP.  
OS Bos taurus.  
FH key Location/Qualifiers  
FT Peptide 1..18  
FT /label= sig\_peptide  
FT Peptide 19..537  
FT /label= mat\_peptide  
PN 007107979-A.  
PD 25-APR-1995.  
PE 15-AUG-1994; 214227.  
PR 19-AUG-1993; JP-227881.  
PA (KINO-) KINOSSEI PEPTIDE KENKYUSHO KK.  
DR WPI: 95-190179/25.  
DR N-PSDB; Q90442.  
PT New DNA encoding an oviduct-specific glycoprotein - useful for  
PT recombinant protein production in high quantities.  
PS Claim 4; Pages 7-9; 22pp; Japanese.  
CC G00442 encodes R73991 bovine oviduct specific glycoprotein (BOGP).  
CC The cDNA can be used for the commercial recombinant prodn. of  
CC BOGP in high quantities.  
SQ Sequence 537 AA;

Query Match 48.9%; Score 1328; DB 14; Length 537;

Best local Similarity 51.2%; Pred. No. 9.77e-114;  
Matches 187; Conservative 74; Mismatches 97; Indels 7; Gaps 6;

Db	20	klycyfnwafsrpgpasllprldpfclthvfafasmsnqlypkdpqekillypefn	79
Qy	2	klycyfntmaoyrgearflpkdpslchlyafagmtnhoistewndet-lyoeffn	60
Db	80	klyerngkltllslgwmvgtyrftmlstfnrervfsvallthtgfdgldlffly	139
Qy	61	GLKMANFKLTLLAIGWNGTQRFIDWVAIANNRQTFVSAIRFLKYSFGDLDLWERY	120
Db	140	pglrsparardwrfvllleellqafkneqlltmprlllaaavgdphvqkayearllg	199
Qy	121	PGSGGSAVDKERTTTLVQDLANAFQEDATQSGKERLLLSAAVPAQGTYYDAGYDVXIA	180

```

Db 200 rllgftswlswdlngswkxvghnspfls-l-pgddks-sa-yamyywqlgvppekl 254
Qy 181 QNDFVNLMAVDFHGSWEKVTGHNSPLTKROESSGAALSLWDAAVQMLQKTPASKLI 240
Db 255 mglptvgrtffhllkasqnelragavspaspkyrlkqagflayaelccfvrrakrwind 314
Qy 241 LGMPYTGSRFLTLASSDPRVGAPRPTGSGTPEPFLKEGMLAYIVYCSW-KGARIKRQRDQ 259
Db 315 yvypafskxewvgyddaisfgyakfllkrehfganmwldldlfsyfcgctpfpvlvt 374
Qy 300 KVPYTFRDNQMWGDDVVSFKTYSYKLQKGLGAMWAWLDDDFAGFSCNOGRYPPIQR 359
Db 375 lnlll 379
Qy 360 LRQEL 364

```

RESULT	2
ID	R73993 standard; Protein; 668 AA.
AC	R73993;
DT	10-JAN-1996 (first entry)
DE	Hamster oviduct specific glycoprotein.
KW	Hamster oviduct specific glycoprotein; recombinant production;
KW	HOGP.
OS	Cricetulus griseus.
PS	Key
FT	Peptide
FT	1..18
FT	/label= sig_peptide
FT	Peptide
FT	19..668
FT	/label= mat_peptide
PN	U07107979-A.
PD	25-APR-1995.
PF	15-AUG-1994; 214227.
PF	19-AUG-1993; JP-227881.
PA	(KINO-) KINOSEI PEPTIDE KENKYUSHO KK.
DR	WP1. 95-190179/25.
DR	N-PSDB: Q90444.
PT	New DNA encoding an oviduct-specific glycoprotein - useful for
PT	recombinant protein production in high quantities.
PS	Claim 4; Pages 16-18; 22pp; Japanese.
CC	O90444 encodes R73993 hamster oviduct specific glycoprotein (HOGP).
CC	The cDNA can be used for the commercial recombinant prodn. of
CC	HOGP in high quantities.
CC	Sequence 668 AA;
CC	Sequence 668 AA;

Query Match	48.0%;	Score 1303;	DB 14;	Length 668;
Best Local Similarity	48.4%;	Pred. No. 2.57e-111;		
Matches 177;	Conservative	81;	Mismatches 99;	Indels 9;
				Gaps 7;

[illegible]

```
Db 374 11nel1 379
    | : |
QY 359 TLRQEL 364
```

ID	RESULT	3
AC	R73992; standard; Protein; 718 AA.	
AC	R73992;	
DE	10-JAN-1996 (first entry)	
DE	Murine oviduct specific glycoprotein.	
KM	Murine oviduct specific glycoprotein; recombinant production;	
KM	MOCp.	
OS	Mus musculus.	
FH	key	Location/Qualifiers
FT	peptide	1..18
FT	/label= sig_peptide	
FT	peptide	19..718
FT	/label= mat_peptide	
PN	J07107979-A.	
PD	25-APR-1995.	
PD	15-AUG-1994; 214227.	
PR	15-AUG-1993; JP-227681.	
PA	(KINO-) KINOSER PEPTIDE KENKYUSHO KK.	
DR	WPI: 95-190179/25.	
DR	N-PSDB; Q90443.	
FT	New DNA encoding an oviduct-specific glycoprotein - useful for	
PT	recombinant protein production in high quantities.	
PS	Claim 4; Pages 11-14; 22pp; Japanese.	
CC	G90443 encodes R73992 murine oviduct specific glycoprotein (MOCp)	
CC	The cDNA can be used for the commercial recombinant prodn. of	
CC	MOCp in high quantities.	
CC	Sequence 718 AA;	
50		

Query Match	48.0%;	Score 1303;	DB 14;	Length 718;
Best Local Similarity	49.9%;	Pred. No. 2.57e-11;		
Matches	182;	Conservative	76;	Mismatches 100;
			Indels	7;
			Gaps	4;

Db	20	kllcyftwahrgrpsimphdi:dpfclchllfaafsmnqglvknldenvlypefn	79
Oy	2	kllvcftfwmAoyRQGEARFLPKDDPSLCTHLIAFAAGMTNHOJSTEWNDEY-LTQEFN	60
Db	80	kIkernelkclllsIgvnfgtsrftamlstlanrekIfidsvlsflrbfdgldlffly	139
Oy	61	GLKKNNPLKTLTLAIGGNFNETQKTLDMVATANNRQGFVNSAIRFLUKYSFDDGLDMEX	120
Db	140	pqlIygsphd:wnflfiIeelqfefe:eaalltqnprlllsaaVsgubslhtsydalllg	199
Oy	121	PGSOSPSVYDEKERTTIVODIANAFQDAQPSGSEKRLLSAAVPAQGYVDAGYEYDKIA	180
Db	200	rll:dfinrlsydlIgswekfctghaspl:fsIpedsksa--y---amIywkIlgtpadkl	254
Oy	131	QNLDPVNLMAIDPFHSWEKTYGHNSP:PLTKROEESGAALSLNWDAAVQOMLQKGTSPASKLI	240
Db	235	mgfptygrnfyllkkesknqgltaemgpaSpqkyLkqgflayevcsIvgrakhhwIdyq	314
Oy	241	LGMPYGRSFTLIASSDSPTRWGAPATGSGTGPPEFKEGMLAYEVECSW-KGATKQRIQDQ	299
Db	315	yypaIfkskewlgyddtIsfykamykrehIfggamwvltldmdddvrgtfgngpfplvhl	374
Oy	300	KVPYIFRNQWVGDDVESEFTKYSYLRKQGLGAGAMWALDLDDFAFGSCNQGRIYPLIQ	359
Db	375	Inell 379	
Oy	360	LQREL 364	
RESULT 4			
AC	P81342	standard: protein; 321 AA.	
DE	P81342:		
DE	19-OCT-1980	(first entry)	
KW	Polypeptide involved in protective mechanisms		
PN	Immune response; cell growth.		
	J63032898-A.		







OY 301 VPIYF-RD-NQWVGFDVSEFKTKVSYLKQKGLGAMW 337

RESULT 9  
ID R33072 standard; Protein: 423 AA.  
AC R33072;  
DT 30-JUN-1993 (first entry)  
DE Aphanocladium album pre-pro-endochitinase.  
KM filamentous fungus; pathogen resistance.  
OS Aphanocladium album.  
FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /label- signal  
FT /note- "pre-sequence"  
FT Peptide 23..34  
FT /note- "pro-sequence"  
FT Protein 35..423  
FT /label- endochitinase  
PN EP-531218-A.  
PD 10-MAR-1993.  
PF 04-SEP-1992; 402414.  
PR 06-SEP-1991; FR-011072.  
PA (SNFI) ELF SANOFI.  
PA (ERAP) SOC NAT ELF AQUITAINE.  
PI Blaiseau P, Legoux R, Leguay J, Schneider M;  
DR WPI: 93-078700/10.  
DR N-PSDB: Q37711, Q37712.  
PT Recombinant DNA encoding fungal endochitinase - useful for  
PT protecting plants against fungi, insects, etc.  
PS Claim 1; Fig 1; 75pp; French.  
CC Clone CH3C was isolated from a cDNA bank prepared from A. album  
CC mRNA by screening with anti-chitinase antiserum. Three probes were  
CC designed based on the CH3C sequence, corresponding to the start,  
CC middle and downstream regions of the coding sequence. The probes  
CC were used to isolate the full-length coding sequence of A. album  
CC endochitinase from an A. album cDNA library in pTZ19R. A 1.6kb  
CC fragment was identified in several positive clones. The sequence  
CC of the fragment contained an ORF encoding a 423 amino acid protein.  
CC The mature endochitinase can be used to control pathogens  
CC such as fungi, bacteria, arthropods and nematodes on plants.  
SQ Sequence 423 AA.

Query Match 21.2%; Score 576; DB 6; Length 423;  
Best Local Similarity 36.2%; Pred. No. 1.01e-41;  
Matches 101; Conservative 65; Mismatches 95; Indels 18; Gaps 16;

DB 115 LKQGRNMKVMVLSIGVWST-NFPAASSAATKTKFAGSAGFMKDWGFGDIDWEYP 173  
OY 62 LKKNPKLTKTLAIGWNGFTQKFTDMVATANNROTFFVNSAIRFLRKYSFSDGLDWEYP 121

DB 174 -adatqagmvlllgavreldsyaay-akg-hhflislaapadpnykikfa-e-1g 228  
OY 122 GSGGSADVADKERTTLVQDLANAFQEAQTSKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

DB 229 kvldynlmaydyagswnytghdaniyanpnunatp-yntddavqayingvpankiv 287  
OY 181 QNLDPEVNLMAVDFHGSMEVYTGHNPSPLYKROESGAAASLNDAVQOMLOKTPASKLI 240

DB 288 lmplygrsfq-qt--eg-1gkpyng1gs-gsw--englwdykalpk-agatvkcdttak 339  
OY 241 LCMPTYGRSFTLASSSDTRVGAPATSGTPPGFTKRGSMALAYEVCSWKGATKQRIQDOK 300

DB 340 gcysydpstckelstfdtpamistkxswlkkgk1gsgmfw 378  
OY 301 VPIYF-RD-NQWVGFDVSEFKTKVSYLKQKGLGAMW 337

RESULT 10  
ID R33068 standard; Protein: 389 AA.  
AC R33068;  
DT 30-JUN-1993 (first entry)  
DE Aphanocladium album mature endochitinase.

KW filamentous fungus; pathogen resistance.  
OS Aphanocladium album.  
PN EP-531218-A.  
PD 10-MAR-1993.  
PF 04-SEP-1992; 402414.  
PR 06-SEP-1991; FR-011072.  
PA (SNFI) ELF SANOFI.  
PA (ERAP) SOC NAT ELF AQUITAINE.  
PI Blaiseau P, Legoux R, Leguay J, Schneider M;  
DR WPI: 93-078700/10.  
DR N-PSDB: Q37704.  
PT Recombinant DNA encoding fungal endochitinase - useful for  
PT protecting plants against fungi, insects, etc.  
PS Claim 22; Page 60-61; 75pp; French.  
CC This mature protein from the filamentous fungus Aphanocladium  
CC album has endochitinase activity and is used to control pathogens  
CC such as fungi, bacteria, arthropods and nematodes on plants.  
CC Recombinant DNA comprising the endochitinase coding sequence is  
CC specifically intended for transforming Nicotiana tabacum, Helianthus  
CC annuus and Brassica napus to pathogen resistance.  
SQ Sequence 389 AA.

Query Match 20.9%; Score 569; DB 6; Length 389;  
Best Local Similarity 35.8%; Pred. No. 4.58e-41;  
Matches 100; Conservative 66; Mismatches 95; Indels 18; Gaps 16;

DB 81 LKQGRNMKVMVLSIGVWST-NFPAASSAATKTKFAGSAGFMKDWGFGDIDWEYP 139  
OY 62 LKKNPKLTKTLAIGWNGFTQKFTDMVATANNROTFFVNSAIRFLRKYSFSDGLDWEYP 121

DB 140 -adatqagmvlllgavreldsyaay-akg-hhflislaapadpnykikfa-e-1g 194  
OY 122 GSGGSADVADKERTTLVQDLANAFQEAQTSKERLLLSAAVPAG-QTYVDAGYEVDKIA 180

DB 195 kvldynlmaydyagswnytghdaniyanpnunatp-yntddavqayingvpankiv 253  
OY 181 QNLDPEVNLMAVDFHGSMEVYTGHNPSPLYKROESGAAASLNDAVQOMLOKTPASKLI 240

DB 254 lmplygrsfq-qt--eg-1gkpyng1gs-gsw--englwdykalpk-agatvkcdttak 305  
OY 241 LCMPTYGRSFTLASSSDTRVGAPATSGTPPGFTKRGSMALAYEVCSWKGATKQRIQDOK 300

DB 306 gcysydpstckelstfdtpamistkxswlkkgk1gsgmfw 344  
OY 301 VPIYF-RD-NQWVGFDVSEFKTKVSYLKQKGLGAMW 337

RESULT 11  
ID W02159 standard; Protein: 866 AA.  
AC W02159;  
DT 14-JAN-1997 (first entry)  
DE Soluble chitinase.  
KM Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;  
KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;  
OS catabolism.  
OS Vibrio furnissii.  
PN W09625424-A1.  
PD 22-AUG-1996.  
PF 13-FEB-1996; U02332.  
PR 13-FEB-1995; US-386727.  
PA (UYIO) UNIV JOHNS HOPKINS.  
PI Bassler B, Chittiaru E, Keyhani N, Roseman S, Rowe C;  
DR WPI: 96-39335/39.  
DR N-PSDB: T36390.  
PT Chitin biosynthetic enzymes and I, exo I and exo II - are  
PT periplasmic chitodextrinase(s), periplasmic beta-GlcNAcidase(s) and  
PT aryl beta-N-acetylglucosaminidase(s), respectively  
Example 4; Page 79-82; 101pp; English.  
CC Periplasmic chitodextrinase (W02156), periplasmic  
CC beta-N-acetylglucosaminidase (W02157) and aryl  
CC beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
CC oligosaccharides with the structure (GlcNAc)n where n is 2 or



[illegible]

RESULT 14  
ID R32547 standard; Protein; 238 AA.  
AC R32547;  
DT 17-JUN-1993 (first entry)  
DE HaNPV IE-1 gene protein (partial sequence).  
KW *Heliothis armigera* nuclear polyhedrosis virus; Ha; NPV;  
KW insecticide; immediate early-1.  
PS *Heliothis armigera* nuclear polyhedrosis virus.  
PN W09303144-A.  
PD 18-FEB-1993.  
PF 05-AUG-1992; AU0413.  
PR 05-AUG-1991; AU-007576.  
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
PI Christian PD, Dail DJ, Gordon KHT, Hanzlik TN, Siskancha A.  
DR WPI: 93-075601/09.  
DR N-PSDB: Q3582.  
DT Insect virus with reduced capacity to occlude viral particles - used  
PT for controlling proliferation of insect pests without horizontal  
PT transmission  
PT Disclosure: Fig 2: 51pp; English.  
CS The sequence is that encoded by the *Heliothis armigera* (Ha)  
CC nuclear polyhedrosis virus (NPV) immediate early-1 (IE-1) gene  
CC (partial sequence).  
SQ Sequence 238 AA;

ID	RESULT
AC	15
AC	W02156 standard; Protein; 1046 AA.
AC	W02156; (first entry)
DE	14-JAN-1997
DE	Periplasmic chitodextrinase.
DE	Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase
KW	Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW	catabolism.
OS	Vibrio furnissii.
PN	W09625424-A1.
PN	22-AUG-1996.
PD	13-FEB-1996; U02332.
PD	13-FEB-1995; US-386727.
PR	(UJJO ) UNIV JOHNS HOPKINS.
PA	Baszler B, Chitlaru E, Keyhani N, Roseman S, Rowe C;
PI	Yu C.
PI	WPI; 96-393335/39.
DR	WPI-PSDB; T36387.
DR	Chitin biosynthetic enzymes end I, exo I and exo II - are

PT periplasmic chitin: dextrinase(s), periplasmic beta-GlcNAc(s) and  
PT aryl beta-N-acetyl-gluco:amidase(s), respectively  
PS Claim 2; Page 68-71; 101pp: English.  
CC Periplasmic chitinodextrinase (W02156), periplasmic  
CC Beta-N-acetylglucosaminidase (W02157) and aryl  
CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin  
CC oligosaccharides with the structure (GlcNAc)<sub>n</sub> where n is 2 or  
CC higher, by contacting them with soluble chitin. The enzymes are  
CC encoded by the genes *exlI*, *exoI* and *exoII* respectively. They are  
CC all genes involved in the catabolic pathway of chitin.  
Q0 Sequence 1046 AA;

Search completed: Fri Jun 27 17:10:12 1997  
Job time : 102 secs.





Oy	62	LKKANPKLTLLAIGMWNFGQKFTDVAANAANNQTFVNSAIRLRKRYSFDGDLDMEYR	121
Dd	143	grr-----dknhfttlkkmkaefikeaq-pqkqklllsalsagkvltidsydiaktsq	196
Oy	122	GSGSPAVDKRRFRTLVDLANAQOEAKTSNGKERLLLSAAVPAGGYVADAGEVVKIAQ	181
Dd	197	hldfslmtvdfbhawgtctghshplfrfgedaaapdrfsntdyavymrl-lgsapaasklv	256
Oy	182	NLDENVNLMAYDFHGSMEKWVTGHNSPLTKROESSGAALSLNVDAVQAOMLKGTPASKTLT	241
Dd	257	giptfgrsfclaae-eelgvgapispgp1pgrftkeagtlayelcldlrgatvrhtlrqg	315
Oy	242	GMPIYGRSFTLIASSSDIRVGAPARGSTPGPFTKEGGMLAYEVCSW-KKATRKORIDOK	300
Dd	316	vpyatkgnqvwgyddqgesvsksvylkdrglagamwvalddtfqgefsgqdltfrpltna	375
Oy	301	VPLYFRDNQMWGFPDVESFKTKRYLYLKQKGLGAMWVALDDLDPAGSCNQG-RYPLIQR	359
Dd	376	lkdaia 381	
Oy	360	LROELIS 365	
RESULT	2		
ENTRY	2	S51327 #type complete	
TITLE		heparin-binding glycoprotein 38k - pig	
ORGANISM		#formal_name Sus scrofa domestica #common.name domestic pig	
DATE		07-May-1995 #sequence_revision 07-May-1995 #text_change 07-May-1995	
ACCESSIONS		S51327	
REFERENCE		S51327	
authors		Shackleton, L.M.; Mann, D.M.; Mills, A.J.T.	
#submission		Submitted to the EMBL Data Library, January 1995	
#description		Identification of a 38kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodelling.	
#accession		S51327	
##status		Preliminary	
##residues		1-383 ##label SHA	
##cross-references		EMBL:Z47603	
SUMMARY		#length 383 #molecular-weight 42443 #checksum 2907	
Query Match		52.6%, Score 1428; DB 16; Length 383;	
Best Local Similarity		52.2%; Pred. No. 4,41e-239;	
Matches 191; Conservative		81; Mismatches 85; Indels 9; Gaps 5;	
Dd	23	klyveyswsagyregdgscfdaiidfliclhlysfanjsnmeidletwmndvcltydtlnt 82	
Oy	2	KLVCFPTNMAGYRGGEARFLPRKDLPStCHLLYAFAFGMTNHOLSTEWMDETLYOEFNG 61	
Dd	83	lknrnpnlktllsvgyvmfgsqrfsklasntqsrrfkasvvpoflrthgfgdglawisp 142	
Oy	62	LKKANPKLTLLAIGMWNFGQKFTDVAANAANNQTFVNSAIRLRKRYSFDGDLDMEYR 121	
Dd	143	grr-----dkxhlttlkemkaefivreal-ptterilllgsvagsakvaIdrgydiasq 196	
Oy	122	GSGSPAVDKRRFRTLVDLANAQOEAKTSNGKERLLLSAAVPAGGYVADAGEVVKIAQ 181	
Dd	197	hldfslmtvdfbhawgtctghshplfrfgedaaasdfsnaadvavsyvl-lgsapanklvm 256	
Oy	182	NLDENVNLMAYDFHGSMEKWVTGHNSPLTKYROESSGAALSLNVDAVQAOMLKGTPASKTLT 241	
Dd	257	giptfgrsfclaae-ftdvgapaspplprftfkegltlayelcldlrgatvrtpirpqg 315	
Oy	242	GMPIYGRSFTLIASSSDIRVGAPARGSTPGPFTKEGGMLAYEVCSW-KKATRKORIDOK 300	
Dd	316	vpyatkgnqvwgyddqesvsnkakakytkrsqjagawmvcltlddfrmgfcgqnrlfpjltsa 375	
Oy	301	VPLYFRDNQMWGFPDVESFKTKRYLYLKQKGLGAMWVALDDLDPAGSCNQG-RYPLIQR 359	
Dd	376	ikdvia 381	

[illegible]



[illegible]

KEYWORDS	1-483 #label KRI	
SUMMARY	glycosidase; hydrolase	
Query Match	43.4%; Score 1179; DB 12; Length 483;	
Best Local Similarity	42.5%; Pred. No. 1,28e-191;	
Matches 158; Conservative	98; Mismatches 99; Indels 17; Gaps 11.	
Db	23 kvvqyfgawsvyrrgngkfddngldpeltchlllysfvngvkgdkvklidpsvldpnlidgf 82	
Qy	2 klvcyftwmaoyrggeaflfepkldpdlcthllyafagmnhhoistye-wnd-ety-y 56	
Db	83 gkftslrkknpsvklmawvgwmaagsvrfpsqmasdgaetrafaqnvklfdqyfgiddf 142	
Qy	57 qefnqlkkmnklktlliaigwmngtqkftfdmvaTANNROTfVNSAIRfLRKISfSDLDL 116	
Db	143 dweyqatrgsgpadvkmvklckalkafvq--h---d-yisaavaapetsasksyd 195	
Qy	117 dMEYPSGSGSPADVdkERETTLVQDLANAFQEQADTSKERLLLSAAVPAGQTYVDAYEV 176	
Db	196 aemsgyldfmltmydflfmgvwdhtgmhappsasbhsnmlklnvkaavkylngvpxk 255	
Qy	177 dkIAQNLDfVWMAKYDFGSGWEKTKGHNSPLRYKQESGSAASLNVDAAVQOMLQKTPA 236	
Db	256 ekIvvgvayvskftlinspnkylgaavsgaqtaqpytngnglllyvleicemkagdwv 315	
Qy	237 SKLLIGMPTYKRSTTLSSSDTRKGAARATSGTGPPTKEGMLAYEVCSW-KGATKQR 295	
Db	316 vqdhkgyvavkngvqswfdlialakakqgflkqegljggawmsletddfkgl-cge-k 373	
Qy	296 IOD-QK-VPYfFRDNOwMGFDVDSFRTKSYLKQKLGAMVWALDLDfAFGSCNQGR 353	
Db	374 ypyvkalnsylvg 385	
Qy	354 YPLIQTLRKQELS 365	
RESULT	9	
ENTRY	A38221	#type complete
TITLE	chitinase (EC 3.2.1.14) MPl - nematode (Brugia malayi)	
ORGANISM	formal_name Brugia malayi	
DATE	04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996	
ACCESSIONS	A38221	
REFERENCE	A38221	
#authors	Fuhman, J.A.; Lane, W.S.; Smith, R.F.; Plessens, W.F.; Perlier, F.B.	
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1548-1552	
#title	Transmission-blocking antibodies recognize microfilarial chitinase in brugian lymphatic filariasis.	
#cross-references	MUDD:92179220	
#accession	A38221	
#status	preliminary; not compared with conceptual translation	
#molecule_type	nucleic acid; protein	
#residues	1-504 #label FvH	
#cross-references	NCBI:85345	
#note	sequence extracted from NCBI backbone	
KEYWORDS	glycosidase; hydrolase	
SUMMARY	length 504 #molecular-weight 55971 #checksum 7343	
Query Match	40.0%; Score 1086; DB 12; Length 504;	
Best Local Similarity	42.8%; Pred. No. 5.49e-174;	
Matches 158; Conservative	94; Mismatches 102; Indels 15; Gaps 11.	
Db	27 cytncwagdydgskflpignlpglchlllyafakvdelgskpfewndetewskgmys 86	
Qy	5 cyfTnwAoyRGGEARfLPKLDpSLCTHLLYAFAGMNHOLSTY-EWNDE-T-----LYQ 57	
Db	87 avtklrenmpkljlylsgygvfsgafstfgakaaacterfiksafalflknmfdgfdl 146	
Qy	58 EFNGLKkMNPfLKLTLIAIGWMNGTQKFTFDVAVANNROTfVNSAIRfLRKISfSDGLD 117	



Db 147 weyr-v-gv-a-e-eh-akiweamktafeeaektsqkrl1llaasagkqtdidsynve 200  
118 WEPGSGSPAVDKERFTTLVQDLANAFQEOATOSKERLLLSAAPAGOTYDAGYEV 177  
Qy 118 WEPGSGSPAVDKERFTTLVQDLANAFQEOATOSKERLLLSAAPAGOTYDAGYEV 177  
Db 201 slgknfdllfimydlhgswekndvthgk1hptkgevsigifntefaaadysagmpke 260  
Qy 178 KIAQNDLVNLMAYDFHSGSEKVTGHNSPYKRQOEESGAASINVDAVQOMLOKGPAS 237  
Db 261 kl1lgfpmayqgtdlnpsetalgaasrpsasaknpagtsyweiclylkegkety 320  
Qy 238 kl1lgfpmayqgtdlnpsetalgaasrpsasaknpagtsyweiclylkegkety 296  
Db 321 hgegyaymwkgdgyqwydneetrlkmw1kkyggaftwafidfdftkscgkqpy 380  
Qy 297 QDQKV-PYIFRDQWQVGFDDVESFKTKVSYLKQKGLGAMVWALDDDFGFCNGGRYP 355  
Db 381 lnaasael 389  
Qy 356 LIQTLROEL 364

RESULT 10  
ENTRY A56596 #type complete  
TITLE chitinase (EC 3.2.1.14) - tobacco hornworm  
ORGANISM #formal\_name Manduca sexta #common\_name tobacco hornworm  
DATE 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 03-May-1996

ACCESSIONS A56596  
REFERENCE A56596  
#authors Kramer, K.J.; Corpuz, L.; Chol, H.K.; Muthukrishnan, S.  
#journal Insect Biochem. Mol. Biol. (1993) 23:691-701  
#title Sequence of a cDNA and expression of the gene encoding epidermal and gut chitinases of Manduca sexta.  
#cross-references M01D:93357793  
#accession A56596  
#status preliminary  
#molecule\_type mRNA  
#residues 1-554 #label KRA  
#cross-references NCBI:136417; NCBI:136418  
#experimental\_source larvae  
#note sequence extracted from NCBI backbone  
KEYWORDS glycosidase; hydrolase  
SUMMARY #length 554 #molecular\_weight 62203 #checksum 4659

Query Match 37.4%; Score 1015; DB 12; Length 554;  
Best Local Similarity 40.4%; Pred. No. 1,44e-160;  
Matches 155; Conservative 98; Mismatches 107; Indels 24; Gaps 18;

Db 23 ar1vcyfsnawayrpygyrlyediprekthlysfygtgensev11ldpeldvdkng 82  
Qy 1 AK1VCFTTMAQRQGEARFLPRDLPSLCTHLYAFAGMTN-H-Q-LST-TEMN-DETL 55  
Db 83 frntfsjrsbhvklfmaavgaesgskysmwagctsmfirsfsf1kkydfqld 142  
Qy 56 YQFENGKAKNNPKTKTLAIGWNNFGTKFTDMATANNKOTVNSAIRLRKXSPGLD 115  
Db 143 ldeyppaadrsgsfcdkilylvgelrra---lrv-gkyvel-taavplnfrlmeq 197  
Qy 116 ldeyppaadrsgsfcdkilylvgelrra---lrv-gkyvel-taavplnfrlmeq 197  
Db 198 yhpelqelidalhmsydlrlnmagfadhprlykxprhphqwyeklnvndghlweeq 257  
Qy 174 YEVDKIAQNDLVNLMAYDFHSGSEKVTGHNSPYKRQOEESGAASINVDAVQOMLOK 233  
Db 258 cspnklvgyrlyfgrsftlaagannylgtfinkeagqgdpaytnatgfwayelctev 317  
Qy 234 TPASKILIGMPTGRSTTLASSDT-RVGA-P-ATGSGTGPFTKGGMLAYEVNS-- 267  
Db 318 dkdsgwtckwdegqkcpaykgtqvwgyedprsvelkmw1kkygylgamtalnddf 377  
Qy 288 WKG-A--TKORIODOKPYIFRDQWQVGFDDVESFKTKVSYLKQKGLGAMVWALDDDF 344  
Db 378 qgl-cge-kp1klk1lhkms-sy 398

Qy 345 AGFSCNCGRPILQTLROELSLPY 368  
Db 11 A28368 #type complete  
ENTRY chitinase (EC 3.2.1.14) precursor - Bacillus circulans  
TITLE #formal\_name Bacillus circulans  
ORGANISM #formal\_name Bacillus circulans  
DATE 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 03-May-1996

ACCESSIONS A28368  
REFERENCE A28368  
#authors Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka, H.  
#journal J. Biol. Chem. (1990) 265:15659-15665  
#title Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed its evolutionary relationship to Serratia chitinase and to the type III homology units of fibronectin.  
#cross-references M01D:90368776  
#accession A28368  
#status preliminary  
#molecule\_type DNA  
#residues 1-699 #label WAT  
#cross-references GB:J05599  
CLASSIFICATION #superfamily fibronectin type III repeat homology  
KEYWORDS glycosidase; hydrolase  
SUMMARY #length 699 #molecular\_weight 73677 #checksum 6757

Query Match 25.7%; Score 698; DB 10; Length 699;  
Best Local Similarity 40.1%; Pred. No. 2,65e-101;  
Matches 119; Conservative 73; Mismatches 85; Indels 20; Gaps 18;

Db 144 qnkllyqtnpnlkllsvgwtw-snrfsdaataatrevfanasavdlrkynfdvld 202  
Qy 58 EFNGLKKNPKTKTLAIGWNNFGTKFTDMATANNKOTVNSAIRLRKXSPGLD 117  
Db 203 weyrvsgldgnskrpedqnyllsklreklidaagavdk-kyl1llasgasaya-a 260  
Qy 118 WEPGSGSPAVDKERFTTLVQDLANAFQEOATOSKERLLLSAAPAGOTYDAGYEV 177  
Db 261 ntelakiaavdwlmtydfngawqksahnaplydpaaagypdantfvaagag 320  
Qy 173 GYVDKIAQNDLVNLMAYDFHSGSEKVTGHNSPYKRQOEESGAASINVDAVQOMLOK 228  
Db 321 hldyppaaklylgyrlyfgrgwd-gcagagn-gqygtctgssvgtweagsfidylean 378  
Qy 229 WLGKGPASKLIGMPTGRSTTLASSDT-RVGA-P-ATGSGTGPFTKGGMLAYEVNS-- 267  
Db 379 ylnkngylyrlyndakvpyllynsnkrflsyddaeavgyktaylkskylgagmfwel 435  
Qy 288 W-K-GATKORIODOKPYIFRDQWQVGFDDVESFKTKVSYLKQKGLGAMVWAL 339

RESULT 12  
ENTRY S51591 #type complete  
TITLE chitinase (EC 3.2.1.14)/lysozyme (EC 3.2.1.17) P2 precursor,  
ORGANISM #formal\_name Nicotiana glauca - common tobacco  
#formal\_name Nicotiana glauca - common tobacco  
DATE 15-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 03-May-1996

ACCESSIONS S51591  
REFERENCE S51591  
#authors Heitz, T.; Second, S.; Kaufmann, S.; Geoffroy, P.; Prasad, V.; Brunner, F.; Fritig, B.; Legendre, M.  
#journal Mol. Gen. Genet. (1994) 245:246-254  
#title Molecular characterization of a novel tobacco pathogenesis-related (PR) protein: a new plant chitinase/lysozyme.  
#accession S51591  
#molecule\_type mRNA  
#residues 1-378 #label HEI  
#cross-references EMBL:X78325

```

#accession S51632
##experimental_source cv. Samsun NN
##molecule_type protein
##residues 3136;87-112;252-275;282-305;337-371 ##label HEW
CLASSIFICATION #superfamily Streptomyces chitinase ch140
KEYWORDS glycosidase; hydrolase
FEATURE
1-25 #domain signal sequence #status predicted #label STG\
26-371 #product chitinase/lysozyme PZ #status predicted #label
371-378 MAT\
SUMMARY #domain carboxyl-terminal propeptide #status
#length 378 #molecular_weight 42019 #checksum 6689

Query Match 21.9%; Score 594; DB 5; Length 378;
Best Local Similarity 32.7%; Pred. No. 3.36e-82;
Matches 106; Conservative 87; Mismatches 108; Indels 23; Gaps 18;

Db 43 nmlstclthfcafadpnsqnlispengds-fagfstvgrknpvxtfisiagtr 101
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 23 KDLDPSLCTHLIYAFAGMT--NHQLSTEWMDLRYLGEFNG-LKKMMPKLTILAT-GGW 78
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 102 a-dttayigimarqpnrsksfidssrlarqfghlaldkeyp--l-s-aldmtnlgll 156
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 79 NFGTKFMDNATANNRQTFVNSAIRFLRKYSFQGLDMDWEPESQSPAVDKRERFTLV 138
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 157 newrtalimeernsgraillltaavsyssprvnglnlyvesvarnlwnlwnaydflypnw 216
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 139 QDLANAFQOEAOISGKERLLLSAIVPAQOTVYDAGEYDKRINQIDLVDVNLMAVDFHGS-W 197
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 217 spsgntshaqfddpnrh--lsgs--d6-anawtgaqvpeklklylfpfygawrlvnpn 270
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 198 EKV-TGHSPLPKROESGAAASLNDVAVOQMLOKGTSPKSLTLGMPYTGSRSTLASSS 256
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 271 lhdrlpaagksnyva-vddgsm-lynnrlrlylvsratvynaatlvgdycysgmwlsy 328
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 257 DTRVGAPATGSGTGPFPFKKEGMLAYEVCWKGATQR-IQDQK-VP-YIFRNMQWVF 313
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 329 ddtgsvrnkvnyvkgrgllygfaw 352
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 314 DVESFRTKVSYLKRGIGGAMWV 337
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 13
ENTRY S47133 #type complete
TITLE chitinase (EC 3.2.1.14) - fungus (Trichoderma harzianum)
ORGANISM #formal_name Trichoderma harzianum
DATE 23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change 10-May-1996
ACCESSIONS S47133
REFERENCE S47133
AUTHORS Carasollo, C.; Gutierrez, A.; Jimenez, B.; van Montagu, M.;
#submission submitted to the EMBL Data Library, June 1994
#description Characterization of ech-42, a Trichoderma harzianum
endochitinase gene expressed during mycoparasitism.
#accession S47133
#status preliminary
#molecule_type DNA
#residues 1-424 #label CAR
#cross-references EMBL:X79381

GENETICS
#introns 48/2; 81/2; 98/1
CLASSIFICATION #superfamily Streptomyces chitinase ch140
KEYWORDS glycosidase; hydrolase
SUMMARY #length 424 #molecular_weight 46298 #checksum 8773

Query Match 21.4%; Score 582; DB 5; Length 424;
Best Local Similarity 36.0%; Pred. No. 5.18e-80;
Matches 102; Conservative 69; Mismatches 95; Indels 17; Gaps 13;
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Qy	57	QENIGKMKMKNKLTLLAIGMNGTQKFDMDVATANNROTFPVNSAIRFLRKRSFSDLDL	116
Db	170	dweyp-addtqatmavlllkeirsqdayaagy-apy-yhllislaapagpehy-sfilhm	225
Qy	117	DMEYSGSGSPAVDKERFFTLVODLANAFQEOAOTSGKERLLLSAAPAGOTYVDAGY	176
Db	226	sdlqgvdlyvnlmaydyagswsyysgldanlfanspmsnp-yntdqakdykqgvpa	284
Qy	177	DKIAQNDLVNLMAYDFHSGWEKVTGHNSPLYRKQOESSGAASLMDVAAVOOMQOKTPA	236
Db	285	skivlgmlyiraef-estqg-lyqytsyqigs-gsw-englwdy-kvlpkagatvyqd	336
Qy	237	SKLLIGMHTYRSTLLASSSDTRRGARATGSGTGPPTKEGMLATYEVCMSKCATQ--	294
Db	337	svagayysdpskskellsfcdpmdntkvsyiknlgjgsmfw	379
Qy	295	RIDOKVPIYIFRDQNWGQFDDVESFETKTVSLKQKGGAGAWV	337
RESULT	14		
ENTRY			
TITLE			
ORGANISM			
DATE			
ACCESSIONS			
REFERENCE			
#authors			
#journal			
#title			
#accession			
##molecule_type			
##residues			
##cross-references			
COMMENT			
GENETICS			
#gene			
#introns			
CLASSIFICATION			
KEYWORDS			
FEATURE			
1-34			
35-423			
SUMMARY			
Query Match	21.2%	Score 576;	DB 5; Length 423;
Best Local Similarity	36.2%;	Pred. No. 6,42e-79;	
Matches 101;	Conservative 65;	Mismatches 95;	Indels 18; Gaps 16;
Db	115	Lkqgrnmkwmllslygvtwst-nfpaassaatrkftagsavgfmmkwdgididweyp	173
Qy	62	LKKANPKKLTLLAIGMNGTQKFDMDVATANNROTFPVNSAIRFLRKYSFSDGLDMEYR	121
Db	174	-adatqagmmvlllqavrseldsyaaqy-ekg-ghfllislaapagpnykikfa-e-lg	228
Qy	122	GSQSPAVDKERFFTLVODLANAFQEOAOTSGKERLLLSAAPAG-QTYVDAGYVDKIA	180
Db	229	kvldylnlmaydyagswsnyqghaaniyaanpnpnatp-yntdavgayingvpankiv	287
Qy	181	QNDLFVNLMAYDFHSGWEKVTGHNSPLYRKQOESSGAASLMDVAAVOOMQOKGPSKLI	240
Db	288	lgmlyirsfg-qt-eg-igkpyngigs-gsw-englwdykalpk-agatvkcddtak	339
Qy	241	LGMFTYGRSFLASSSDTRVGAPATGSGTGPPTKEGMLATYEVCMSKCATQRIODCK	300
Db	340	gcysydsptkelisfdtpmstksvslkqjlgsmfw	378
Qy	301	VPIYIF-RD-NQWGFDDVESFETKTVSLKQKGGAGAWV	337

```

RESULT      15
ENTRY       S51369      #type complete
TITLE       Chitinase - fungus (Trichoderma harzianum)
ORGANISM    #formal_name Trichoderma harzianum
DATE        15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change
            03-May-1996
ACCESSIONS  S51369
REFERENCE   S51369
#authors    Garcia, I.; Lora, J.M.; de la Cruz, J.; Benitez, T.; Llobell,
            A.; Pinfor-Toro, J.A.
#journal    Curr. Genet. (1994) 27:83-89
#title      Cloning and characterization of a chitinase (CHIT42) cDNA
            from the mycoparasitic fungus Trichoderma harzianum.
#accession  S51369
#status     Preliminary
#molecule_type DNA
#residues   1-423 #label GAR
SUMMARY     #length 423 #molecular_weight 46056 #checksum 2000

Query Match      20.6%; Score 561; DB 12; Length 423;
Best Local Similarity 35.7%; Pred. No. 3,43e-76;
Matches 101; Conservative 71; Mismatches 94; Indels 17; Gaps 13;

Db      111 qlfk-vkkanrglkvlstgwtst-nfpaastdanrknfaktaitfmkgfdgidi 168
QY      57 QEFNGLKKNPKLKTLLATIGWNFGIOKFTDMVATANNROTENVSAIRFLRKISFDGLD 116

Db      169 dweyp-adatqaasm111kevsqrdayaay-apy-yhfl11laapagkdy-sk1r1 224
QY      117 DWYEPGSGSPAVDKERFTLVQDLANAFQEAQTSKERLLLSAAVPAGOTYVDAGYEV 176

Db      225 adlgyvldylnlaydagsfsplctghdanlfmnpnpatp-fntdsavkdylingvpa 283
QY      177 DKIAQNIDFVNLAAYDFHGSMEKVTGHNSPLYKROESGAASLNVDAAYQOWLOKGTPA 236

Db      284 nkiivgmprygsf--qntag--igqlyngvgs-gsw--eaqlwdykalpk-agatvqyd 335
QY      237 SKLILGMPRTYGRSFTLASSSDTRYGAPATSGTPGPTKEGMLAYTEVCWKGATKORI 296

Db      336 svakgyysnsackelsfdtpdminkvaylkslqigsfmfw 378
QY      297 QDQKVPIYFRD--NQWVGFPDVESFKTKVSYLKNKGIGAMWV 337

```

Search completed: Fri Jun 27 17:06:56 1997  
 Job time : 134 secs.



\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
Mparch\_p protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 27 17:07:13 1997; MasPar time 12.20 Seconds  
Tabular output not generated.  
648.449 Million cell updates/sec  
\*\*\*\*\*

Title: >US-08-663-618A-15  
Description: (1-373) from US08663618A.pep  
Perfect Score: 2717  
Sequence: 1 AKVCYFTNMAQYRQGEARF.....YPLIQLTRELSELPYLSGT 373  
Scoring table: PAM 150  
Gap 11  
Searched: 59021 segs, 21210388 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 48.971; Variance 85.460; scale 0.573

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1488	54.8	383	4	GP39_HUMAN CARTILAGE GLYCOPROTEI	0.00e+00
2	1086	40.0	504	2	CHIT_BDUA ENDOCHITINASE PRECURS	1.03e-218
3	1015	37.4	554	2	CHIT_MANSE ENDOCHITINASE PRECURS	7.72e-102
4	698	25.7	699	2	CHIT_BACCI CHITINASE A1 PRECURSO	2.43e-127
5	576	21.2	423	2	CHIT1_APHAL CHITINASE 1 PRECURSOR	3.13e-99
6	561	20.6	423	2	CHIT4_TRIHA 42 KD ENDOCHITINASE P	8.30e-96
7	546	20.1	427	2	CHIL_COCIM ENDOCHITINASE 1 PREC	2.16e-92
8	437	16.1	563	2	CHIA_SERVA CHITINASE A PRECURSOR	7.14e-68
9	395	14.5	551	2	CHIT_NPVAC PROBABLE ENDOCHITIN	1.33e-58
10	345	12.7	820	2	CHIA_ALTSO CHITINASE A PRECURSOR	9.66e-48
11	337	12.4	499	2	CHIB_SERVA CHITINASE B PRECURSOR	5.03e-46
12	301	11.1	619	2	CHIT_STRLI CHITINASE C PRECURSOR	2.19e-38
13	287	10.6	610	2	CHIT_STRLI CHITINASE 63 PRECURSO	1.86e-35
14	263	9.7	1146	5	KTXA_KIUDA KILLER TOXIN ALPHA AN	1.67e-30
15	172	6.3	36	4	GP39_BOVIN CARTILAGE GLYCOPROTEI	8.53e-13
16	172	6.3	83	3	EDOS_PAPAN ESTRADIOL-DEPENDENT O	8.53e-13
17	121	4.5	597	2	CHIX_STRCI EXOCHITINASE I PREC	3.64e-04
18	108	4.0	452	4	HEMF_RHOSH COPROPORPHYRINOGEN II	3.11e-02
19	110	4.0	497	2	CHID_BACCI CHITINASE D PRECURSOR	1.60e-02
20	104	3.8	422	10	UMUC_ECOCI DMIC PROTEIN.	1.14e-01
21	100	3.7	364	11	YBR5_YEAST HYPOTHETICAL 40.9 KD	4.02e-01
22	101	3.7	1045	7	PRTT_SERMA EXTRACELLULAR SERINE	2.94e-01

23	101	3.7	1070	1	AGIU_CANTS ALPHA-GLUCOSIDASE PRE	2.94e-01
24	99	3.6	398	9	SPEB_STRPY EXOTOXIN TYPE B PRECU	5.48e-01
25	98	3.6	492	9	SYK_THERH LYSYL-TRNA SYNTHETASE	7.44e-01
26	97	3.6	563	5	K2CB_HUMAN KERATIN, TYPE II CYTO	1.01e+00
27	97	3.6	563	5	K2CA_HUMAN KERATIN, TYPE II CYTO	1.01e+00
28	97	3.6	563	5	K2CC_HUMAN KERATIN, TYPE II CYTO	1.01e+00
29	99	3.6	958	1	AMYG_SCHOC GLUCAMYLASE 1 PREC	5.48e-01
30	98	3.6	1045	7	PRTS_SERMA EXTRACELLULAR SERINE	7.44e-01
31	96	3.5	297	5	HXK1_HUMAN HOMEBOX PROTEIN MSX-	1.36e+00
32	94	3.5	344	4	GCH2_PICGU GTP CYCLOHYDROLASE II	2.47e+00
33	95	3.5	353	11	YOH2_BACSU PUTATIVE PEPTIDASE IN	1.84e+00
34	94	3.5	404	6	MOEA_HAEIN MOLYBDOPEROXIN BIOSYNT	2.47e+00
35	95	3.5	554	5	HYES_RAT SOLUBLE EPOXIDE HYDRO	1.36e+00
36	96	3.5	554	5	HYES_RAT SOLUBLE EPOXIDE HYDRO	1.36e+00
37	95	3.5	563	5	K2CE_HUMAN KERATIN, TYPE II CYTO	1.84e+00
38	95	3.5	563	5	K2CF_HUMAN KERATIN, TYPE II CYTO	1.84e+00
39	95	3.5	563	5	K2CG_HUMAN KERATIN, TYPE II CYTO	1.84e+00
40	94	3.5	993	11	YAJ1_SCHPO PUTATIVE FAMILY 31 GL	2.47e+00
41	95	3.5	1333	10	VGRL_MOUSE VASCULAR ENDOTHELIAL	1.84e+00
42	93	3.4	267	3	EBAG_FILASP ENDO-BETA-N-ACETYLGLU	3.31e+00
43	93	3.4	295	11	YH08_YEAST HYPOTHETICAL 35.1 KD	3.31e+00
44	92	3.4	510	1	BMP6_MOUSE BONE MORPHOGENETIC PR	4.43e+00
45	92	3.4	557	3	ERK3_HUMAN EXTRACELLULAR SIGNAL-	4.43e+00

## ALIGNMENTS

RESULT 1  
ID GP39\_HUMAN STANDARD; PRT; 383 AA.  
AC P36222; P30923;  
DT 01-JUL-1993 (REL. 26, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE CARTILAGE GLYCOPROTEIN-39 PRECURSOR (GP-39) (39 KD SYNOVIAL PROTEIN).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX TISSUE-CARTILAGE;  
RX MEDLINE: 9406458.  
RA HAKALA B.E., WHITE C., RECKLIES A.D.;  
RL J. BIOL. CHEM. 268:25803-25810(1993).  
RN [2]  
RP SEQUENCE OF 22-45.  
RX MEDLINE: 90328983.  
RA NYIKOS P., GOLDS E.E.;  
RL BIOCHEM. J. 269:265-268(1990).  
CC - FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE CAPACITY OF CELLS TO  
CC RESPOND TO AND COPE WITH CHANGES IN THEIR ENVIRONMENT.  
CC - TISSUE SPECIFICITY: PRESENT IN ARTICULAR CHONDROCYTES, SYNOVIAL  
CC CELLS AS WELL AS IN LIVER. UNDETECTABLE IN MUSCLE TISSUES, LUNG,  
CC PANCREAS, MONONUCLEAR CELLS, OR FIBROBLASTS.  
CC - PTM: GLYCOSYLATED.  
CC - SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC - SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.  
DR EMBL: M80927; G348912; -  
DR PIR: S10677; S10677.  
DR PIR: A33162; A33162.  
DR PROSITE: PS01095; CHITINASE\_18.  
KW GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 22 383  
FT CARBOHYD 60 60  
SQ SEQUENCE 383 AA: 42613 MW, D0A5280F CMC32;  
Query Match 54.88; Score 1488; DB 4; Length 383;  
Best Local Similarity 53.38; Pred. No. 0.00e+00;  
Matches 195; Conservative 83; Mismatches 79; Indels 9; Gaps 5;  
Db 23 klvcyysvsgyregdscfpdldrlfchlllyfsfnisnhdtkvndvlygmnt 82  
OY 2 klvcyftnmaoyrgearflpdlpslcthllyafagkgnholstewndetlyofeng 61

```

Db      83 lkntrpnukllsvgywnfgsqrfsklasntqsrtrffikswprflrthgddglawlyp 142
Qy      62 LKRNMPKRLTLAAGWMEFGQKQFTDMVATNNQNTFVNSAIRLRXFSFGDLDMEYp 121
Db      143 grf-----dkghfttlkkmkaefikeag-pgkqklllaaslsqkvkltssydyaktsq 136
Qy      122 GSQGSFPAVDKRFRTTLVODLNNAPQOEKQVSGKRELLLSAIVPAGQYVVDGHEVDKIAQ 181
Db      197 hldfslmtcyfhwargwtgetheqplfrgqgsedafsdntdyvvymlrjgaspasklvm 256
Qy      182 NLDFNNMAYDFHGSWMEVNTGHNSPLTKYRQESSAASLNNDAVQOMLKGTPASKTLIL 241
Db      257 giprfgrsfclass-etvgarjisgprjprgfckeaqtlayaelcldflrgatvhrtlrgq 315
Qy      242 GMPTRGSRFTLASSSDTRFVGAPATGSGPBPFTREGMLAYEVCNW-KGATRGRIQDQK 300
Db      316 vpyattkgnqwgvgddqesvksvqylkdrqlagamvvalldldfqsfcgqdlfrpfltna 375
Qy      301 VPYIFRDMQWNGFDDVESFRTKVSYLAKQGLGGMVVALDLDPAFGFSCNQG-RYPLIQF 359
Db      376 lkdaa 381
Qy      360 LKQELS 365

```

RESULT	2	STANDARD:	PRT:	504 AA.
ID	CHIT. BROMA			
AC	P29030:			
DT	01-DEC-1992 (REL. 24, CREATED)			
DT	01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	ENDOCHITINASE PRECURSOR (EC 3.2.1.14) (MEL ANTIGEN).			
OC	BUGIA MALAYI.			
SC	EURAROTIA, METAPOA, ACCELLOMAIES, NEMATODA, SECCEREMENTEA, SPIRURIDA.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE; 92119220.			
RA	FURMAN J.A., LANE W.S., SMITH R.F., PIESSENS W.F., PERLER F.B.;			
RL	PROC. NATL. ACAD. SCI. U.S.A. 89:1548-1552(1992).			
CC	-1- FUNCTION: THE MEL ANTIGEN IS A MICROFILARIAE CHITINASE, WHICH MAY			
CC	FUNCTION TO DEGRADE CHITIN-CONTAINING STRUCTURES IN THE MICRO-			
CC	FIARIA OR IN ITS MOSQUITO VECTOR DURING PARASITE DEVELOPMENT AND			
CC	TRANSITION.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF			
CC	N-ACERYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-1- PTM: O-GLUCOSYLATED.			
CC	-1- KNOWN TO BIND CALCIUM.			
CC	-1- DEVELOPMENTAL STAGE: THE APPEARANCE OF THE MEL ANTIGEN CORRESPOND			
CC	WITH THE ONSET OF THE PARASITE'S ABILITY TO INFECT THE MOSQUITO.			
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES).			
DR	EMBL; M73689; G156064; -.			
DR	PIR; A38221; A38221.			
KR	PROSITE: PS01095: CHITINASE_18.			
KM	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ANTIGEN;			
FW	REPEAT; GLYCOPROTEIN; CALCIUM-BINDING.			
FT	SIGNAL	1	22	
FT	CHAIN	23	504	ENDOCHITINASE.
FT	DOMAIN	23	400	CATALYTIC.
FT	DOMAIN	401	450	SER/THR-RICH (LINKER).
FT	DOMAIN	407	448	3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT	ACT_SITE	148	148	PROTON DONOR (BY SIMILARITY).
QO	SEQUENCE	504 AA;	55971 MW;	4DAY5EB CRC32;

[illegible]

```

Db 87 avtklirempg\kxlllygygrnfsaifcylaksaqkterfiksalafrknmfdgfd 146
Qy 58 EFNGKXKMPKIKTLKLLAIGGNEGTQFTMVMVATANNRQTFVSAIRFLERKYSFSDGHD 117
Db 147 weyo-v-gv-a-e-eh-akiveamktafveeaksgkqrllltaeasagqtlldsyvne 200
Qy 118 WEIPGSOCSPAVNDERFTTLVODLANNFQOEAOQSGKERLLLSAAVAPAGQTYVDAGIEVD 177
Db 201 slgkncdillmsydlhngsweknndlhgkllhpckgevsygiintefaaadyscmpe 260
Qy 178 KIAQNLDPVNLMAADDFGSMWKYTGHSPLXKROEBGAAASLNVDAVQOMLQKGPAS 237
Db 261 kiliqimpyagqwtldhpsetaigaasrpsasaktnpagtsayweicklykkgkety 320
Qy 238 KLIMFMYGSRFTLASSSDTRVGCAPATGSGTPEPFRKEGGMLAYVECSW-KGATQRI 296
Db 321 hgevgayamvkgddwydydneetlrlkmkylkekgyggaftwaldfddfgskscgkypp 380
Qy 297 QODKRV-PIFRDNONOVGFDDVESEFKTVSYLKKOKGLOGAGVAMVALLDDDFEGSCNOGRYP 355
Db 381 llnaiskel 389
Qy 356 LIQTLROEL 364

```

RESULT	3	STANDARD;	PRT;	554 AA.
ID	CHIT_1MANSE			
AC	P36362;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	ENDOCHITINASE PRECURSOR (EC 3.2.1.14)			
OS	MANDUCA SEXTA (TOBACCO HAWMOTH) (TOBACCO HORNWORM).			
OC	EDUKAROTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 93357793.			
RA	KRAMER K.J., CORPUZ L., CHOI H.K., MUTHUKRISHNAN S.;			
RL	INSECT BIOCHEM. MOL. BIOL. 23:651-701(1993).			
CC	-1- FUNCTION: DIGEST CHITIN IN THE EXOSKELETON DURING THE MOLTING PROCESS.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF			
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.			
CC	-1- DEVELOPMENTAL STAGE: HIGH LEVELS SEEN IN THE EPIDERMIS ON DAY 0,			
CC	BUT RAPIDLY DISAPPEARS AND IS UNDETECTED ON DAYS 1-4 OF FIFTH			
CC	INSTAR. IT REAPPEARS ON DAY 5 AND PEAKS ON DAY 7 AFTER WHICH A			
CC	RAPID DECLINE IS SEEN. IN THE GUT IS DETECTED ON DAY 6 WITH LOWER			
CC	LEVELS SEEN ON DAYS 0, 7 AND 8.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS AND GUT.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES).			
CC	EMBL: U02270; G406049; -;			
DR	EMBL; S64577; -; NOT_ANNOTATED CDS.			
DR	HSSP; P07254; ICTN.			
DR	PROSITE: PS01095; CHITINASE_18.			
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	554	ENDOCHITINASE.
FT	DOMAIN	396	453	SER/THR-RICH.
FT	ACT_SITE	146	146	PROTON DONOR (BY SIMILARITY).
FT	CARBOHYD	85	85	POTENTIAL.
FT	CARBOHYD	303	303	POTENTIAL.
FT	CARBOHYD	407	407	POTENTIAL.
FT	CARBOHYD	545	545	POTENTIAL.
SO	SEQUENCE	554 AA;	62203 MW;	FA87F8AD CRC32;

Query Match	37.4%	Score 1015;	DB 2;	length 554;
Best Local Similarity	40.4%;	Pred. No. 7.72e-202;		
Matches	155;	Conservative	98;	Mismatches 107; Indels 24; Gaps 18;
Db	23	ariveyfnwavyppgyrryiediprekchillysflytregnewllldpeldvknk	82	

Matches	119;	Conservative	73;	Mismatches	85;	Indels	20;	Gaps	18
Db	144	qlnkjktgnpnlktllisvgwtw-snifsdvaataatrevfansavgdflrkynfdgvdld	202						
Qy	58	EFNGIKKNPKTKTLTLLAGGWNFGTQKFTDVAATANNQTFVNSAIRLRKXSFEDGLDLD	117						
Db	203	weypsggldgnskrpekknytl1llk1trkilaagaavgk-kyllttiasgaaty-a	260						
Qy	118	WEYP-GS-QG-S-PAYDKERFTLLVODLANAFQOEQTSKEKRLLSAAVPAQOTYVDA	172						
Db	261	nteljakiaetdwnimtydfengaqwk1sahnap1ndpaaasaavgpdantfvaagag	320						
Qy	173	GVEVDK1QNLDPEFNALAYDFHGSNEKVTGNISFL-YKROE-ESGAA-A-SLNDAAVQO	228						
Db	321	hldavgpaak1vjlgvpfygrgdw-gcaqaqn-ggyqctcggssvgtveasgfdfydlean	378						
Qy	229	WLQKGTIPASK1LLGMPYTRGFSFLTIASSSDTRVGAAPATGSGTGPFGFTREGGLAYEY-CS	287						
Db	379	ylnkngytrryndactkvpvlylnasnkrfisyddaeavgktyak1kskyljggamfwe1	435						
Qy	288	W--K-GATKOR1ODQKVPYIFRD-NQW-VGDDVDSEFTKVSYLKQGLGAGAWMAL	339						
RESULT	5	STANDARD;	PRT;	423	AA.				
ID	CH11	APHAL							
AC	P32470;								
DT	01-OCT-1993	(REL. 27, CREATED)							
DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)							
DT	01-FEB-1996	(REL. 33, LAST ANNOTATION UPDATE)							
DE	CHITINASE 1	PRECURSOR (EC 3.2.1.14).							
GN	CH11.								
OC	APPANOC1ADIUM ALBUM.								
OC	EUFAROTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN-ETHM 483;								
RX	MEDLINE; 93013040.								
RA	BLAISEAU P.-L., LAFAÏ J.-F.;								
RL	GENE 120:243-248 (1992).								
RN	[2]								
RP	SEQUENCE OF 35-57.								
RX	MEDLINE; 92136437.								
RA	BLAISEAU P.-L., GRISON R., BERTHEAU Y., BRYGOO Y.;								
CC	CUBR. GENET. 21:61-66(1992).								
CC	-1- CATALACTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF								
CC	N-ACETYL-D-D-GLYCOSAMINE POLYMERS OF CHITIN.								
CC	-1- SUBCELLULAR LOCATION: SECRETED.								
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL								
CC	HYDROLASES).								
CC	EMBL; X64104; G429026; -.								
DR	PIR; J01975; J01975.								
DR	HSSP; P07254; ICTN.								
DR	PROSITE; PS01095; CHITINASE_18.								
FM	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN.								
FT	SIGNAL	1	22	POTENTIAL.					
FT	PROPEP	23	34	POTENTIAL.					
FT	CHAIN	35	423	CHITINASE 1.					
FT	ACT SITE	171	171	PROTON DONOR (BY SIMILARITY).					
SO	SEQUENCE	423	AA;	46058	MM;	88456E6E	CRC32;		
Query Match		21.2%;	Score 576;	DB 2;	Length 423;				
Best Local Similarity		36.2%;	Pred. No. 3.13e-99;						
Matches	101;	Conservative	65;	Mismatches	95;	Indels	18;	Gaps	16
Db	115	lkkgnrmkmkmlslsgyvtwst-nfpaassaatktftagsavgfmkdwgfdgididweyp	173						
Qy	62	LKKMKPKTKTLTLLAGGWNFGTQKFTDVAATANNQTFVNSAIRLRKXSFEDGLDWEYP	121						
Db	174	-adatqagmwtlllgavseldeyaay-akg-bhflslsaagpndnyk1kfa-e-1g	228						
Qy	122	GSQGSFAYDKERFTTLYVODLANAFQOEQTSGRKRLLSAAVPAWG-QTYADAGLEVDKIA	180						
Db	229	kvljdylmdaydyswensytcghdaniyanpqpnaptr-yntdavgaylngvpankiv	287						

```
OY      181 QNLDFVNLMAVDHFSWKEVTGHNSPFLKROEEGSAASLNVDAAYQOMLOKGTPARKLI 240
OY      268 lmpdygrsfq-qt--eg-lkxpynglfs-gsw--engldykalpk-agatvcdctak 339
OY      241 LMPDYGRSFTLASSSDIRVAPATGSCSTPPFPFKEGGLMAYVEVCSWKGTKORIDOK 300
Db      340 gcysydpstcklfscfdpmastkyrswlkxglgdtmfw 378
OY      301 VPIIF-RD-NOMVGCHDDVESRKTAKSYLKOGKGLGAMNW 337
```

ID	RESULT	6	CH14_TRIHA	STANDARD;	PRT;	423 AA.
AC	P48827:					
DT	01-FEB-1996	(REL. 33,	CREATED)			
DT	01-FEB-1996	(REL. 33,	LAST SEQUENCE UPDATE)			
DT	01-FEB-1996	(REL. 33,	LAST ANNOTATION UPDATE)			
DE	42 KD ENDOCHITININASE	PRECURSOR	(EC 3.2.1.14).			
GN	CHIT42.					
OS	TRICHODERMA HARZIANUM.					
OC	EURYOTA; FUNGI; DEUTEROMYCOTINA	(IMPERFECT FUNGI).				
RN	[1]					
RP	SEQUENCE FROM N.A.,	AND SEQUENCE OF 35-52; 93-107; 371-385 & 397-414.				
RX	MEDICINE; 95269313.					
RA	GARCIA I., LORA J.M., LA CRUZ J., BENITEZ T., LOBELLE A.,					
RA	PINTOR-TORO J.A.;					
RL	CURR. GENET. 27:83-89(1994).					
CC	-1- FUNCTION: MORPHOGENETIC ROLE DURING APICAL GROWTH, CELL DIVISION					
CC	AND DIFFERENTIATION (CELL WALL MORPHOGENESIS). ANTIFUNGAL					
CC	AGENT.					
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF					
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.					
CC	-1- SUBCELLULAR LOCATION: SECRETED.					
CC	-1- INDUCTION: SPECIFICALLY INDUCED BY CHITIN AND IS CATABOLITE					
CC	REPRESENTED.					
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL					
CC	HYDROLASES)					
DR	EMBL; S78423; G999376; -					
KM	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; ZYMOGEN;					
KM	CHITIN-BINDING.					
FT	SIGNAL	1	22	POTENTIAL.		
FT	PROPEP	23	34			
FT	CHAIN	35	423	42 KD ENDOCHITINASE.		
FT	ACT SITE	171	171	PROTON DONOR (BY SIMILARITY).		
FT	CARBOND	218	218	POTENTIAL.		
SO	SEQUENCE	423 AA;	46056 MW;	899DA50A CRC32;		

	Query Match	20.6%	Score 561;	DB 2;	Length 423;
	Best Local Similarity	35.7%	Pred. No.	8,30e-96;	
	Matches	101; Conservative	71;	Mismatches	94; Indels 17; Gaps 13;
Dd	111 glfk-vkkanrglkvllslgswtst-nfpaastdanrnfaktaaitfmkdwfgidit	168			
Oy	57 OEFNILKKNPKLKLTLAIGNNEFTQAFMTVMATANRRQFEVSAIRFLKRYSFDDIDL	116			
Dd	169 dweyp-adatqgsnmllilkewsgdrdaeyaqy-apq-yhlilaaapagdqny-sklrl	224			
Oy	117 DMEYEGSOGSPAVDKERTTIVQDLANFQQEAGTSKERLLLSAAVPAGCTYYDAQEY	176			
Dd	225 adlgvldylnlmaydygssfspltyghdaanlfmnpnpatp-fntdsavkdyingsva	283			
Oy	177 DKIAONLDPEVNIMAADDFGSMWEKVGHNSPLYKRDEEGGAASLVNPAAYOQMLOKGTQA	236			
Dd	284 nklivgmpldygrsf-qhtga--lqctynygys-gsw--eaegldwykalpk-agatagyad	335			
Oy	237 SKLILGMPTGYRSLFLASSSDPRVGAPARTGGTGPEPFRKEGMLAYEVCSMKATIKORI	296			
Dd	336 svakgyysnatckelislfdcpdmintkvaylkslglgsmfw	378			
Oy	297 ODOKPVLYFRD--NQWGFDDVESEKTVSYLKOKGLGAAWW	337			

ID	RESULT	7	STANDARD;	PRT;	427 AA.
AC	CHILI COCIN				
AD	P54196;				
AE	01-OCT-1996	(REL. 34, CREATED)			
AF	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)			
AG	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)			
AH	ENDOCHITINASE 1	PRECURSOR (EC 3.2.1.14) (COMPLEMENT-FIXATION ANTIGEN)			
AI	(CF-AANTIGEN)	(CF-AG).			
AM	CTSL				
AN	COCCIDIODES IMMITIS.				
AO	EUCHARIOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).				
AP	[1]				
AQ	SEQUENCE FROM N.A.				
AR	STRAIN-C735;				
AS	MEDLINE: 96144270.				
AT	PISHKO E.J., KIRKLAND T.N., COLE G.T.;				
AV	GENE 167:173-177(1995).				
AW	[2]				
AX	SEQUENCE FROM N.A.				
AY	STRAIN-SILVERA;				
AZ	YANG C., ZHU Y., MAGEE D.M., COX R.A.;				
BA	SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBI DATA BANKS.				
BB	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF				
BC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.				
BD	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
BE	HYDROLASES).				
BF	EMBL; L41663; G1200190; -				
BG	DR EMBL; U51271; G1256769; -				
BH	DR EMBL; U33265; G1255728; -				
BI	HYDROLASE. GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;				
BJ	GLYCOPROTEIN.				
BK	SIGNAL	?			
BL	CHAIN	?	427		POTENTIAL. ENDOCHITINASE 1.
BM	CARBOHYD	387	387		POTENTIAL.
BN	CONFLICT	15	47		POTENTIAL.
BO					SMASMNYPVPEAEDEGGFRSVYIVNM (1IN REF. 2).
BP	CONFLICT	199	199		K -> N (1IN REF. 2).
BQ	SEQUENCE	427 AA;	47629 MM;	6CBAAE73	CRC32;

[illegible]



[illegible]

	RESULT	9	STANDARD;	PRT;	551 AA.
AC	CHIT.NPVAC				
DT	P41684;	(REL. 32, CREATED)			
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)			
DE	PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).				
OS	AUTOGRAPH CALIFORNICA NUCLEAR POLYEDROSIS VIRUS (ACMPV).				
CC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-C6;				
RX	MEDLINE; 94303173.				
RA	AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERRER M., POSSEE R.D.;				
RL	VIROLOGY 202:586-605(1994).				
CC	-I- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BEHA-1,4 LINKAGES OF				
CC	N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.				
CC	-I- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
CC	HYDROLASES).				
DR	EMBL; L22858; G559195; -.				
DR	PROSITE; PS00014; ER_TARGET.				
DR	PROSITE; PS01095; CHITINASE_18.				
KW	HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; SIGNAL; GLYCOPROTEIN.				
FT	SIGNAL	1	17	POTENTIAL.	
FT	CHAIN	18	551	PROBABLE ENDOCHITINASE.	
FT	ACT SITE	305	305	PROTON DONOR (BY SIMILARITY).	
FT	CARBONHD	173	173	POTENTIAL.	
FT	CARBOND	444	444	POTENTIAL.	
FT	SITE	548	551	PREVENT SECRETION FROM ER (POTENTIAL).	
SO	SEQUENCE	551 AA;	61368 MW;	BDB124DB CRC32;	
	Query Match	14.5%;	Score 395;	DB 2;	Length 551;
	Best Local Similarity	29.2%;	Pred. No. 1.33e-58;		
	Matches	95;	Conservative	80;	Mismatches 122; Indels 28; Gaps 27;
Dd	221 pwaavqkpkqkvswaw-e-p'yknfgfqjmaaklanphklkipsigwclsdp-fymhdv	277			
Oy	36 AFAGWTNHQLSTENDETLYQ-EFNGLK-KM-NPLKTLTLAGGNNFGOKTDWVAT	91			
Dd	278 -ekrnvfdsykefignwkffdygdicdwefpgygkgnpsjgdaadgcktyllleelram	336			
Oy	92 ANNROTFFNSAIRFLRKRF-DGIDLMEWEGYSOGS-PAV-D-K-E-R-FTTLVODLAMA	144			
Dd	337 lddileac-grv-yeltsaisagdklavnya-e-aqsaigkfllmsydfkgawstldl	392			
Oy	145 FOQ-FAQSSEGERLLLSAAMPAG-QTVVDACYEWDKTAQNLDPFNLAAYPHGSWEVT-	201			
Dd	393 gytqtvyva-ps-vnseelytthyavdalikgvdpnkllivganysgrgtvtntyndy	450			
Oy	202 GHNSPLYRKGESGAASLANVDAVAQGMLOKGTAFSAKLILGMPHYSGSLTIASSSDIRVG	261			
Dd	451 fsgtgng-psgsqtwedgyvdrydqkdlnnyvitfdaagaasyfdkskdqlisfsdvds	509			
Oy	262 APATESGPGEPTFKGGMLAYEVCWSWKATQRIDQ-KVPYIF-RDN-QMWGFDDVES	318			
Dd	510 vlgkyvdrinklglgfaweildan	534			
Oy	319 FKTKVSYLRKQGLGAGMYALDLDD	343			
ID	RESULT	10			
AC	CHIA.ALTSO	STANDARD;	PRT;	820 AA.	
DT	P32823;				
DT	01-OCT-1993	(REL. 27, CREATED)			
DT	01-OCT-1993	(REL. 27, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995	(REL. 31, LAST ANNOTATION UPDATE)			
DE	CHITINASE A PRECURSOR (EC 3.2.1.14) (CHI-A).				
GN	CHIA.				
OS	ALTEROMONS SP. (STRAIN O-7).				
OC	PROKARIOTA; GRACILICTUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;				
OC	VIBRONACEAE.				
RN	[1]				



Db 425 gsdegk-ldaa-dygeaskyldwvymtydfgagwak-ngptaphspitaydgiptggfn 481  
 QY 163 VPAGQTVVDAGYEDVKIAQNLDFVNLMAYDFHSGMEKVTGNSPLKROESGAAA-SLN 221  
 Db 482 tadamakfkakypacklllgfygrgw-vgtsap-ggatgpat-gty--eaagied 536  
 QY 222 VDAVQOMLQKGFPAKSLILGMPYGRSFTLASSSDTRVGA PATGSGTGPFTKEGMLA 281  
 Db 537 y-kvlnscpatgtlag--tayahcgnwswydtpatlksmdwaegqg19gaftwefsg 593  
 QY 282 YVECSMKGATKRIODOKVPIFRDQVWGFFDVESFKTKVSYLKQKGLGAMVWALDL 341  
 Db 594 dtang 598  
 QY 342 DDFAG 346

RESULT 13  
 ID CHIT\_STREL STANDARD; PRT; 610 AA.  
 AC P11220;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE CHITINASE 63 PRECURSOR (EC 3.2.1.14).  
 GN CHTA.  
 OS STREPTOMYCES PLICATUS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92192480.  
 RA ROBBINS P.W., OVERBYE K., ALBRIGHT C., BENFIELD B., PERO J.;  
 RL GENE 111:69-76(1992).  
 RN [2]  
 RP SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.  
 RX MEDLINE: 88087127.  
 RA ROBBINS P.W., ALBRIGHT C., BENFIELD B.;  
 RL J. BIOL. CHEM. 263:443-447(1988).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE BETA-1,4 LINKAGES OF  
 CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
 CC -1- INDUCTION: BY CHITIN.  
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
 CC (CBD).  
 DR EMBL: M82804; G153216; -;  
 DR EMBL: M18397; G153209; -;  
 DR PIR: A29912; A29912.  
 DR PIR: JH0573; JH0573.  
 DR HSSP: P07254; 1CTN.  
 DR PROSITE: PS00018; EF\_HAND.  
 DR PROSITE: PS00561; CBD\_BACTERIAL.  
 DR PROSITE: PS01095; CHITINASE\_18.  
 KM HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL.  
 FT SIGNAL 1 30  
 FT CHAIN 31 610 CHITINASE 63.  
 FT DOMAIN 35 140 CELLULOSE-BINDING.  
 FT DOMAIN 148 229 FIBRONECTIN TYPE-III.  
 FT DOMAIN 236 610 CATALYTIC.  
 FT ACT\_SITE 383 383 PROTON DONOR (BY SIMILARITY).  
 FT CONFLICT 3 3 F -> I (IN REF. 2).  
 SQ SEQUENCE 610 AA; 63974 MW; 2F5EBE35 CRC32;

Query Match 10.6%; Score 287; DB 2; Length 610;  
 Best Local Similarity 26.4%; Pred. No. 1,86e-35;  
 Matches 78; Conservative 63; Mismatches 134; Indels 21; Gaps 16;

Db 320 fnglrnkkaeyphklllysfgrtwsgg-fpdavknpaafakschdlvedprwadvtg1 378  
 QY 56 YQFNGLKKMFKLKTLLAIGGMNGQKFDVATANNRQTFVNSAIRLKRYS-IDGL 114  
 Db 379 dlweyngaglsdcetsapnafsmmkamraefgdyllta-a---vtadgsdgk-id 433

QY 115 DLDWEYSGSG-S--PAVDKERFTTLVQDLANFQQAQTSKGERLLLSAAVPAQGYVD 171  
 Db 434 aa-dygeaskyldwvymtydfgagwak-ngptaphspitaydgiptggftadamakfk 491  
 QY 172 AGEVDKRIQNLDFVNLMAYDFHSGMEKVTGNSPLKROESGAAA-SLNVDAVQOML 230  
 Db 492 skvpacklllgfygrgw-vgtsap-ggatgpat-gty--eaagiedy-kvlnsc 545  
 QY 231 QKGFPAKSLILGMPYGRSFTLASSSDTRVGA PATGSGTGPFTKEGMLAYVECSWK 290  
 Db 546 patgvgag--tayahcgnwswydtpatlksmdwaegqg19gaftwefsgdttng 599  
 QY 291 ATKRIODOKVPIFRDQVWGFFDVESFKTKVSYLKQKGLGAMVWALDLDFAG 346

RESULT 14  
 ID KTXA\_KLUDA STANDARD; PRT; 1146 AA.  
 AC P09805;  
 DT 01-MAR-1989 (REL. 10, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DE KILLER TOXIN ALPHA AND BETA SUBUNITS PRECURSOR (RF2 PROTEIN)  
 DE (ENDOCHITINASE (EC 3.2.1.14)).  
 OS KUUYEROMYCES LACTIS (YEAST).  
 GN PLASMID PGKL-1.  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CBS 2359/152;  
 RA SOR F., FUKUHARA H.;  
 RL CURR. GENET. 9:147-155(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84297209.  
 RA STARK M.J.R., MILLHAM A.J., ROMANOS M.A., BOYD A.;  
 RL NUCLEIC ACIDS RES. 12:6011-6030(1984).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85037931.  
 RA TOKUNAGA M., KAWAMURA A., HISHINUMA F.;  
 RL NUCLEIC ACIDS RES. 12:7581-7597(1984).  
 RN [4]  
 RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.  
 RX MEDLINE: 87004569.  
 RA STARK M.J.R., BOYD A.;  
 RL EMBO J. 5:1995-2002(1986).  
 RN [5]  
 RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.  
 RX MEDLINE: 90259069.  
 RA BRADSHAW H.D. JR.;  
 RL NATURE 345:299-299(1990).  
 RN [6]  
 RP CHITINASE ACTIVITY OF ALPHA-SUBUNIT.  
 RX MEDLINE: 91301161.  
 RA BUTLER A.R., O'DONNELL R.W., MARTIN V.J., GOODAY G.W., STARK M.J.R.;  
 RL EUR. J. BIOCHEM. 199:483-488(1991).  
 CC -1- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH  
 CC THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE  
 CC TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE  
 CC TOXIN) TO GAIN ENTRY INTO THE CELL.  
 CC -1- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-  
 CC SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.  
 CC -1- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,  
 CC BETA AND GAMMA.  
 CC -1- SIMILARITY: LOCAL TO OTHER CHITIN-BINDING PROTEINS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.  
 DR EMBL: X07127; G2829; -;  
 DR EMBL: X00762; G2844; -;  
 DR EMBL: X01095; G2849; -;  
 DR PIR: S07915; S07915.  
 DR HSSP: P02877; 1HEV.  
 DR PROSITE: PS00026; CHITIN-BINDING.

